2007

Investigating reactions catalyzed by terpene synthases in a novel model system

Paul Ross Wilderman
Iowa State University

Investigating reactions catalyzed by terpene synthases in a novel model system

by

Paul Ross Wilderman

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:
Reuben Peters, Major Professor
Alan Myers
Gustavo MacIntosh
Michael Shogren-Knaak
Nicola Pohl

Iowa State University

Ames, Iowa

2007

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I dedicate this thesis to the memory of Paul Lloyd Wilderman,

my father, who was always there for me.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Schemes

Abbreviations

aa	amino acid
AgAS	*Abies grandis* abietadiene synthase
An2	*ent*-CPP synthase from corn
AS	abietadiene synthase
AtKS	*Arabidopsis thaliana* kaurene synthase
cM	centimorgan
CPP	copalyl diphosphate, also called labdadienyl diphosphate
CPS	copalyl diphosphate synthase
cv.	cultivar
DMAPP	dimethylallyl diphosphate
DTS	diterpene synthase
DXP	1-deoxyxylulose-5-phosphate
E. coli	*Escherichia Coli*
FID	flame ionization detection
FPP	(*E*,*E*)-farnesyl diphosphate
GC	gas chromatography
GC-MS	gas chromatography-mass spectrometry
GGPP	(*E*,*E*,*E*)-geranylgeranyl diphosphate
GGPS	(*E*,*E*,*E*)-geranylgeranyl diphosphate synthase
GPP	(*E*)-geranyl diphosphate
GST	glutathione-*S*-transferase
i	indicates indica subspecies of *Oryza sativa*
IPP	isopentenyl diphosphate
j	indicates japonica subspecies of *Oryza sativa*
k_{cat}	catalytic constant

K_M	Michaelis constant
KO(-L)	kaurene oxidase (-like)
KS(-L)	kaurene synthase (-like)
MeJA	methyl jasmonate
MS	mass spectrometry
MVA	mevalonic acid
ORF	open reading frame
OsCPS	*Oryza sativa* copalyl diphosphate synthase
OsCPS1/2	*Oryza sativa ent*-copalyl diphosphate synthases 1 and 2
OsCPS4	*Oryza sativa syn*-copalyl diphosphate synthase
OsDTC1	*Oryza sativa ent*-cassa-12,15-diene diterpene cyclase
OsDTS2	*Oryza sativa* 9β-pimara-7,15-diene diterpene synthase
OsKS1	*Oryza sativa* kaurene synthase
OsKSL	*Oryza sativa* kaurene synthase-like
PaIPS	*Picea abies* isopimara-7,15-diene synthase
PaLAS	*Picea abies* levopimaradiene/abietadiene synthase
r	indicates a recombinant form of the enzyme (i.e.-rAgAS, rAtKS, etc.)
RT	retention time
RT-PCR	reverse transcription-polymerase chain reaction
ssp.	subspecies

Abstract

Terpene synthases are a set of enzymes that initiate the biosynthetic production of the largest class of known natural products. However, not much is known about how these enzymes carry out their complex cyclization reactions to form polycyclic olefins from linear hydrocarbon diphosphates such as geranylgeranyl diphosphate.

Since rice is known to produce many diterpenoid natural products and the rice genome was recently published, sequence information for many putative diterpene synthases was available. This provided the basis for functional characterization of the diterpene synthases from rice (*Oryza sativa* ssp. Indica). An enzyme producing *syn*-pimaradiene was identified, followed by those producing *syn*-stemodene, *ent*-kaurene, *ent*-isokaurene, *ent*-sandaracopimaradiene, *ent*-cassadiene, and *syn*-stemarene. At the same time another group published similar results, however one of the reported *ent*-isokaurene synthases was reported as *ent*-pimaradiene synthase by another research group.

Noting this difference in product with only the three amino acid differences in the active site, experiments were carried out leading to discovery of a single amino acid residue that switches product profile completely to pimaradienes from kaurenes, OsKSL5i:I664T. This switch was found not only in the originally targeted isokaurene synthase, OsKSL5, but also in the second reported isokaurene synthase from rice and kaurene synthases from rice and *Arabidopsis*. Further expounding on this idea, a similar amino acid difference was noted in diterpene synthases of conifers. This amino acid difference, A723S in abietadiene synthase, was a switch, too, with the product profile switching from >95% abietanes to >95% pimaranes.

Utilizing a combination of functional genomics, metabolic engineering, macromolecular modeling, and enzyme biochemistry a basis for further investigations into how terpene synthases work has been provided.

Chapter I. General Introduction

Introduction

The largest class of known natural products, isoprenoids, also known as terpenes or terpenoids, exhibits a wide range of chemical structures and biological activities. This class of compounds has about 50,000 unique members [1]. Labdane-related diterpenes, generally described as being formed from the twenty carbon geranylgeranyl diphosphate (GGPP) and containing the bicyclic core structure found in the labdane family of terpenoids (~7,000 members), are involved in a multitude of biological processes, from growth signaling to defensive measures to pathogenesis [2]. In addition to the role these compounds play in plant physiology, a number of terpenoids are of pharmacological importance.

Terpenes are defined as hydrocarbons of biological origin having carbon skeletons formally derived from isoprene, and terpenoids are natural products and related compounds formally derived from isoprene units [3]. While the structures of terpenes are quite diverse, there are relatively simple linear precursors for these intricate molecules. Synthesis of these precursors initially proceeds via one of two synthesis pathways in plants: the mevalonic acid synthesis pathway, found in the cytosol and provides dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) for the production of farnesyl diphosphate (FPP), the precursor for formation of sesquiterpenes (C_{15}), or the 1-deoxyxylulose-5-phosphate (DXP) pathway, found in plastids providing DMAPP and IPP for the production of monoterpenes (C_{10}) and diterpenes (C_{20}). Terpene synthesis continues with condensation of one molecule of DMAPP and one molecule of IPP in a head-to-tail manner forming GPP, the precursor of monoterpene synthesis; this type of reaction is catalyzed by prenyltransferase enzymes [2]. Addition of another molecule of IPP in a similar head-to-tail reaction produces FPP for sesquiterpene production. Adding one more molecule of IPP provides the GGPP precursor for synthesis of all diterpenes.

These linear precursors are utilized by terpene synthases (TPS), enzymes which further modify the linear GPP, farnesyl diphosphate (FPP), or GGPP. These enzymes provide a basis for the diversity in terpenoids found in nature and are often called cyclases due to the nature of the product formed by the enzyme [2]. These enzymes catalyze complex chemical reactions via a series of a carbocationic intermediates, and product outcome is directed by substrate conformation and active site chemical environment [4]. The initial carbocation is formed one of two ways, via removal of the pyrophosphate group (Class I synthases) or protonation of a C=C double bond (Class II synthases) [5].

Class I and Class II synthases have similar domain structures. Prototypical Class I TPS enzymes have a two domain structure containing regions termed the "N-terminal domain" and the "C-terminal domain" (Figure 1). However, plant TPS enzymes, both Class I and Class II, contain another region of additional N-terminal sequence originally termed the 'insertional' element. In plant TPS enzymes the three regions are the Class IIa, Class IIb, and Class I domains corresponding to the 'insertional', N-terminal, and C-terminal domains, respectively [4-6].

Despite having similar domain structures, Class I and Class II TPS enzymes utilize different active sites and active site architectures [7]. Class I TPS contain a DDXXD aspartate-rich motif that binds divalent metal ions, usually Mg2+, used to facilitate the characteristic allylic diphosphate ionization reaction [4]. This motif is found in the Class I domain of Class I enzymes. A typical Class I TPS catalyzed reaction is that carried out by AtKS (Scheme 1). After removal of the pyrophosphate moiety from *ent*-CPP, the proposed mechanism includes a carbocation shift from C15 to C13. The double bond of C8(17) then attacks the carbocation creating the *ent*-pimaren-8-yl carbocation. A fourth ring is formed by C16 attacking the α face of the *ent*-pimaren-8-yl carbocation. Ring rearrangement occurs when C12 attacks the *ent*-beyeran-16-yl carbocation, which is could be stabilized by the pyrophosphate present in the active site similar to that seen in abietadiene synthase [8]. An

active site base would then remove a proton from C17 of the *ent*-kauren-16-yl carbocation forming *ent*-kaur-16-ene, the product of AtKS.

Class II synthases utilize a spatially separate active site employing a DXDD aspartate-rich motif to catalyze the protonation of a carbon-carbon double bond [7,9]. The catalytic role of the DDXXD motif in Class I enzymes is strongly supported, but the role of the Class II DXDD motif is still being determined [10].

Labdane-related diterpenes are formed in two sequential steps via a Class II TPS and a Class I TPS. Examples of this type of synthesis may be found in gibberellin biosynthesis in flowering plants and in resin acid biosynthesis in conifers (Scheme 2). In gibberellin biosynthesis, the enzymes are copalyl diphosphate synthase (CPS) and kaurene synthase (KS); for resin acid biosynthesis, a bifunctional enzyme, abietadiene synthase (AS), carries out these reactions [11,12].

CPS and KS were two of the first TPS enzymes isolated and biochemically characterized due to the physiological importance of the resulting gibberellin phytohormome [13-16]. However, further studies were limited due to the scarce supply of native enzyme in plants. Recent advances in molecular biology allowed the cloning of KS enzymes from multiple plant species: *Arabidopsis thaliana*, pumpkin, *Stevia rebaudiana* and rice [17-20]. However, full length kaurene synthases contain a plastidial targeting sequence that interferes with proper expression and/or function, and successful expression requires construction of a "pseudomature" form of the protein via truncation of the targeting sequence [9,21-23].

Despite the importance of the biosynthetic machinery for constructing terpenoids, not much is know about the mechanisms by which these products are formed or any of the determinants of substrate specificity or product profile. This state of affairs limits our ability to fully utilize the potential of these natural products.

Recently, there has been significant interest in engineering metabolic pathways of natural products biosynthesis. Much of the interest focused on modular polyketide synthases

and non-ribosomal peptide synthetases. However, terpene synthases are an untapped area of metabolic engineering with intriguing potential, since they are involved in the biosynthesis ~40% of all known natural products [1, 24]. A wide range of hydrocarbon skeletons are formed through a series of carbocation intermediates directed by both the initial folding of the substrate within the active site and enzymatic direction of specific intermediates. This series of events can produce a large number of products, as seen with the known diterpenoid natural products from rice (Scheme 3) [2]. While the substrate may adopt multiple conformations due to large amounts of freedom of rotation within the active site, such variable product outcome severely complicates interpretation of mutational alterations [25]. Nonetheless, product outcome is largely influenced by conformational constraints induced by the active site to fold the substrate. Despite the obvious importance of these influences in engineering product outcome, these phenomena are not well understood. Thus, a system for comparison of TPS producing single products that are highly similar would assist in the analysis of product outcome.

Many plant species produce labdane-related diterpenoids through a single biosynthetic pathway, that of gibberellic acid metabolism [17,18]. However, comparison of product production requires a system containing multiple enzymes utilizing the same substrate and producing different products. This requirement was met when the rice genome was sequenced, and multiple putative TPS genes were found, since rice produces a large number of diterpenoids from the single precursor GGPP (Scheme 4) [26]. With this system, the examination of determinants of product profile could proceed.

The observed selectivity for single stereoisomers of CPP requires tight binding of the bicyclic ring moiety, so differences in substrate and product specificity require significant alterations of the active site cavity to accommodate changes in bicyclic ring conformation and relative positioning of the double bonds due to folding. The recently sequenced rice genome provides a starting point for further studies of Class I TPS enzymes. Knowledge of

determinants of TPS function confers the ability to rationally alter the specificity of DTS enzymes. With the wide conservation across the terpene synthase enzyme family, this work has led to key insights into structure/function relationships in other terpene synthases, too.

Dissertation Organization

This thesis is composed of seven chapters. Chapter I contains an introduction to the topic covered in this thesis and a general outline of how this thesis is structured. Chapters II and III detail functional genomics of TPS from rice (*Oryza sativa ssp. indica*). Chapter IV presents development of a metabolic engineering scheme for facile production of labdane-related diterpenes. Chapters V and VI describe a single residue switch found in Class I TPS enzymes from *O. sativa* and conifers, respectively. Chapter VII details conclusions from this body of work and presents a number of future experiments based on this work.

Chapter II describes the cloning, gene expression pattern, sequence analysis, and biochemical characterization of a *syn*-pimara-7,15-diene synthase from rice as a part of a rice functional genomics project in our lab. This study represented the first portion of my thesis project and also isolated the synthase responsible for producing the carbon backbone for momilactones A & B. It was my first attempt at cloning, analysis and functional characterization of an enzyme. I received a great deal of help with the molecular biology from Meimei Xu, our lab manager and resident molecular biology guru.

As further research was completed on the rice functional genomics project in our lab, we wanted to produce a document describing the entire OsKSL family. The resulting publication is Chapter III. While not the primary author of the study, I performed a large portion of the biochemical characterization of OsKSLs. I also did a significant portion of the sequence analysis for the gene family.

Running concurrently with the research on OsKSLs described in Chapter III, a means of producing larger quantities of product from DTS enzymes was sought. The research on this subject is described in Chapter IV. This metabolic engineering scheme is a powerful means of producing enough product for identification of unknown samples whether they are found singly or are a major or minor component of a mixture. This system was used to assist in the analyses described in Chapters V and VI.

When our laboratory published our examination of the OsKSL gene family, a competing laboratory also published findings on the OsKSL gene family. Much of the research described our publication was supported by the analysis of this other publication [27]. However, one enzyme described in both papers, OsKSL5, produced different products depending on whether it was found in ssp. *indica*, from our research, or ssp. *japonica*, as described by our competitors. After analysis of the amino acid sequence of these two enzymes, a small number of differences were found that mapped to the active site. Mutations were made to change each residue of interest in OsKSL5i to the corresponding residue found in OsKSL5j and vice versa. A single residue was found to change the product profile of OsKSL5i to that of OsKSL5j. This single residue switch was also found to be present in the known KSL enzymes found in rice and also the KS enzymes from pumpkin (*Cucurbita maxima*) and *Arabidopsis thaliana*. This switch "short circuits" the carbocation cyclization/rearrangement cascade of these enzymes. The paper which comprises Chapter V describes the results of these experiments. I performed the initial mutagenesis of OsKSL5i and carried out initial kinetic analysis of AtKS and the mutant AtKS:I638T. I also did a good deal of product analysis of mutants and wild-type enzymes for the OsKSL family enzymes described in this work.

Upon amino acid analysis of the enzymes examined in Chapter V along with the known conifer DTS enzymes, we proposed a similar switch residue for the conifer DTS enzymes. This switch also "short circuits" the carbocation cascade of recombinant

abietadiene synthase from grand fir (*Abies grandis*, rAgAS). The research was performed by me, analysis of the data was done by me and Reuben J. Peters, my advisor, and the paper was written by me and Reuben. This work has been submitted for publication and is in the review process.

References

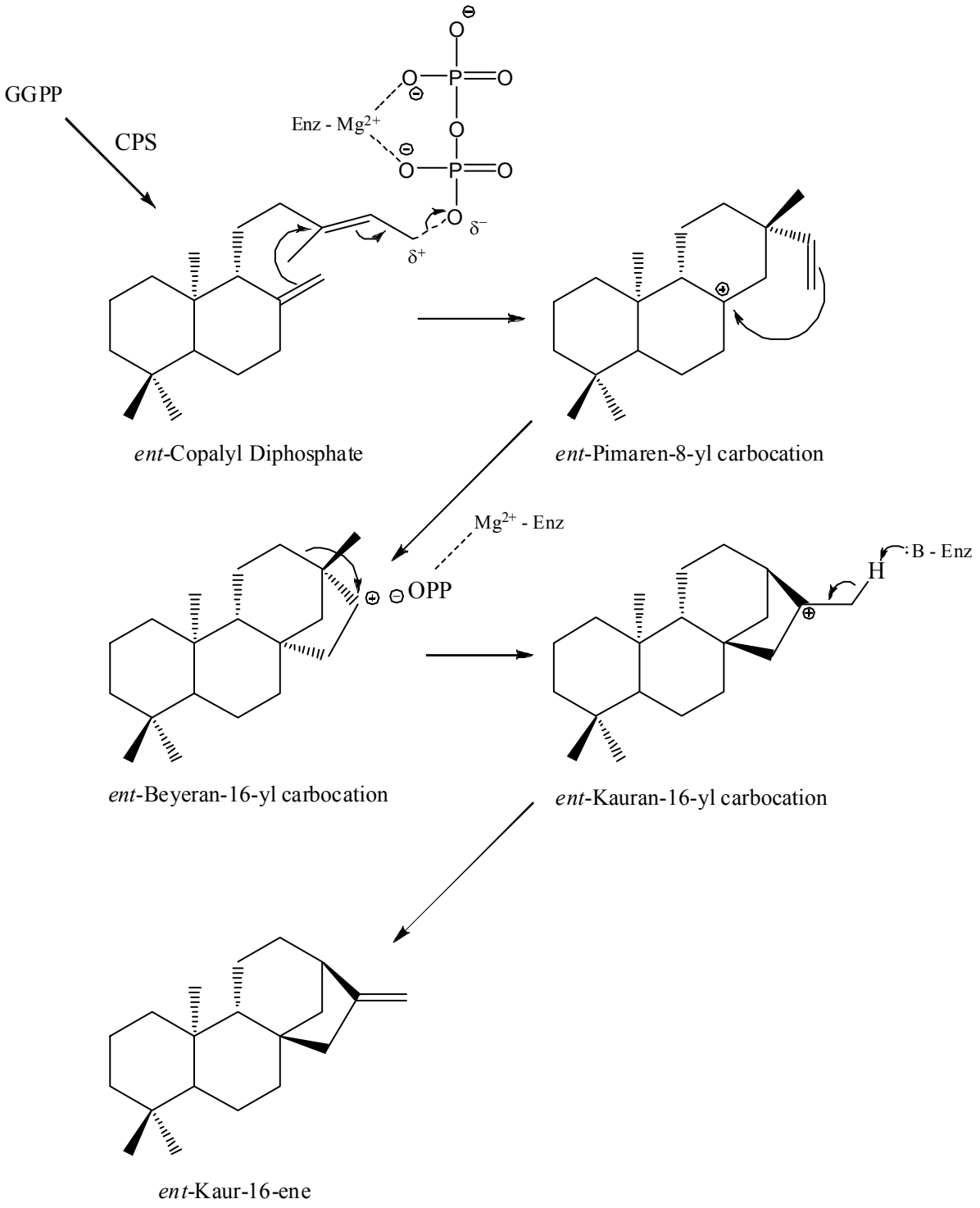
1. Buckingham, J. 2002. Dictionary of Natural Products (on-line web edition). Chapman & Hall/CRC Press.
2. Davis, E.M., and Croteau, R. Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes. *Top. Curr. Chem*., 2000. 209: 53-95.
3. Moss, G.P., Smith, P.A.S., and Tavernier, D. Glossary of Class Names of Organic Compounds and Reactive Intermediates Based on Structure. *Pure Appl. Chem.*, 1995. 67(8/9): 1307-1375.
4. Christianson, D.W. Structural Biology and Chemistry of the Terpenoid Cyclases. *Chem. Rev*., 2006. 106(8): 3412-3442.
5. Wendt, K.U., and Schulz, G.E. Isoprenoid Biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure*, 1998. 6(2): 127-133.
6. Xu, M., Wilderman, P. R., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N., Coates, R. M., & Peters, R. J. Functional Characterization of the Rice Kaurene Synthase-Like Gene Family. *Phytochemistry,* 2007. 68: 312-326.
7. Peters, R.J., Ravn, M.M., Coates, R.M., Croteau, R.B. Bifunctional abietadiene synthase: free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. *J. Am. Chem. Soc.,* 2001. 123: 8974–8978.
8. Ravn, M.M., Peters, R.J., Coates, R.M., and Croteau, R. Mechanism of abietadiene synthase catalysis: stereochemistry and stabilization of the cryptic pimarenyl carbocation intermediates. *J Am Chem Soc*. 2002 Jun 19;124(24):6998-7006.

9. Peters, R.J., Flory, J.E., Jetter, R., Ravn, M.M., Lee, H.J., Coates, R.M., and Croteau, R.B.. Abietadiene Synthase from Grand Fir (Abies grandis): Characterization and Mechanism of Action of the "Pseudomature" Recombinant Enzyme. *Biochemistry*, 2000. 39(50): 15592-15602.
10. Prisic, S., Xu, J., Coates, R.M., and Peters, R.J. Probing the role of the DXDD motif in Class II diterpene cyclases. *Chembiochem,* 2007. 8(8):869-74.
11. Duncan, J.D., and West, C.A. Properties of Kaurene Synthetase from *Marah macrocarpus* Endosperm: Evidence for the Participation of Separate but Interacting Enzymes. *Plant Physiol*., 1981. 68: 1128-1134.
12. LaFever, R.E., Vogel, B.S., and Croteau, R. Diterpenoid Resin Acid Biosynthesis in Conifers: enzymatic cyclization of geranylgeranyl pyrophosphate to abietadiene, the precursor of abietic acid. *Arch Biochem Biophys*., 1994. 313(1): 139-149.
13. Graebe, J.E., Dennis, D.T., Upper, C.D., and West, C.A. Biosynthesis of gibberellins. I. The biosynthesis of (-)-kaurene, (-)-kauren-19-ol, and trans-geranylgeraniol in endosperm nucellus of *Echinocystis macrocarpa*. *J. Biol. Chem.*, 1965. 240(4): 1847-1854.
14. Upper, C.D, and West, C.A. Biosynthesis of gibberellins. II. Enzymic cyclization of geranylgeranyl pyrophosphate to kaurene. *J. Biol. Chem.*, 1967. 242(14): 3285-3292.
15. Fall, R.R., and West, C.A. Purification and Properties of Kaurene Synthetase from *Fusarium moniliforme*. *J. Biol. Chem.*, 1971. 246(22): 6913-6928.
16. Frost, R.G., and West, C.A. Properties of kaurene synthase from *M. macrocarpus*. *Plant Physiol*., 1977. 59: 22-29.
17. Yamaguchi, S., Sun, T., Kawaide, H., Kamiya, Y. The GA2 locus of Arabidopsis thaliana encodes ent-kaurene synthase of gibberellin biosynthesis. *Plant Physiol*., 1998. 116(4): 1271-1278.

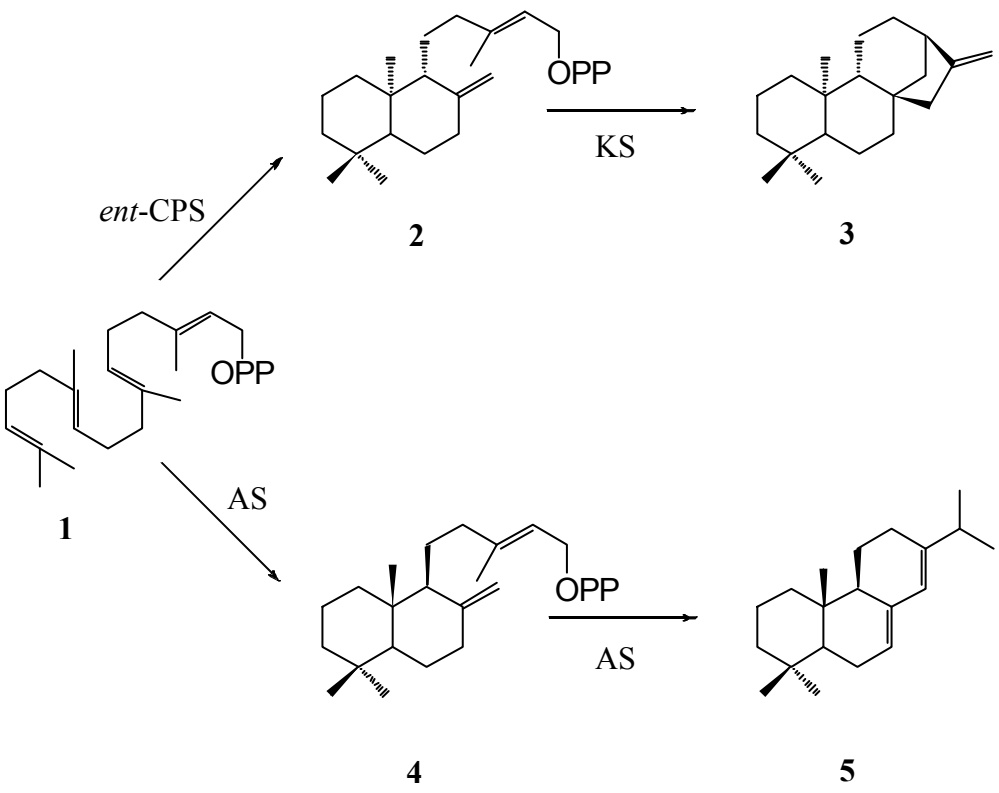
18. Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N., and Kamiya, Y. Molecular cloning and characterization of a cDNA encoding the gibberellin biosynthetic enzyme ent-kaurene synthase B from pumpkin (Cucurbita maxima L.). *Plant J.*, 1996. 10(2): 203-213.
19. Richman, A.S., Gijzen, M., Starratt, A.N., Yang, Z., and Brandle, J.E. Diterpene synthesis in Stevia rebaudiana: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway. *Plant J.*, 1999. 19(4): 411-421.
20. Sakamoto, T., Miura, K., Itoh, H., Tatsumi, T., Ueguchi-Tanaka, M., Ishiyama, K., Kobayashi, M., Agrawal, G.K., Takeda, S., Abe, K., Miyao, A., Hirochika, H., Kitano, H., Ashikari, M., and Matsuoka, M. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol.*, 2004. 134(4): 1642-1653.
21. Bohlmann, J., Meyer-Gauen, G., and Croteau, R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. Proc Natl Acad Sci U S A., 1998. 95(8): 4126-4133.
22. Smith, M.W., Yamaguchi, S., Ait-Ali, T., and Kamiya, Y. The first step of gibberellin biosynthesis in pumpkin is catalyzed by at least two copalyl diphosphate synthases encoded by differentially regulated genes. *Plant Physiol.*, 1998. 118(4): 1411-1419.
23. Williams, D.C., McGarvey, D.J., Katahira, E.J., and Croteau, R. Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry*, 1998. 37(35): 12213-12220.
24. Reeves, C.D. The Enzymology of Combinatorial Biosynthesis. *Crit. Rev. Biotechnol.*, 2003. 23(2): 95-147.
25. Little, D. and R. Croteau, Alteration of product formation by directed mutagenesis and truncation of the multiple-product sesquiterpene synthases δ-selinene synthase and γ-humulene synthase. *Arch. Biochem. Biophys.*, 2002. 402(1): 120-135.

26. Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). *Science*, 2002. 296(5565): 79-92.

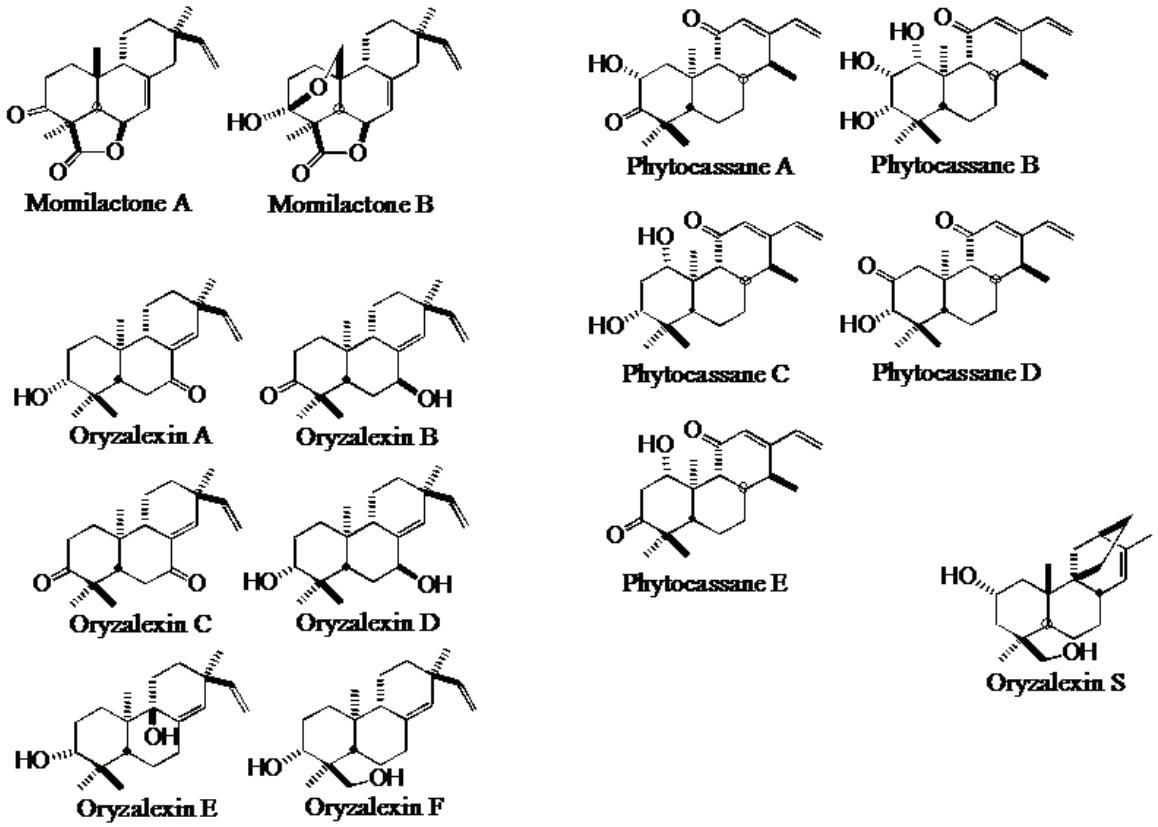
27. Kanno, Y., Otomo, K., Kenmoku, H., Mitsuhashi, W., Yamane, H., Oikawa, H., Toshima, H., Matsuoka, M., Sassa, T., & Toyomasu, T. Characterization of a rice gene family encoding type-A diterpene cyclases. *Biosci. Biotechnol. Biochem*, 2006. 70: 1702-1710.



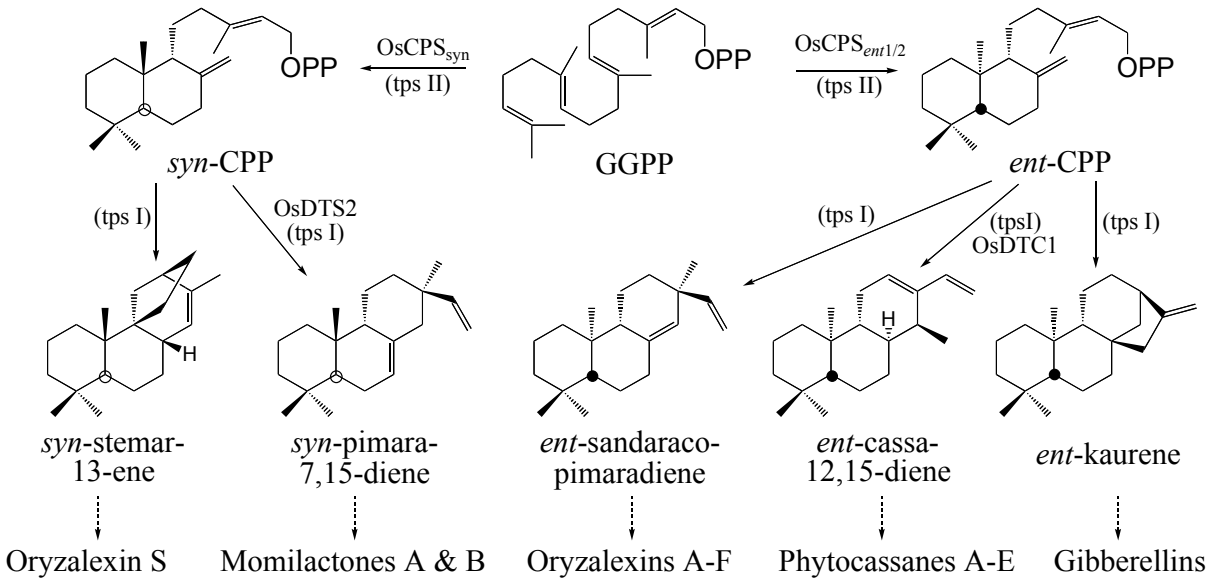
Scheme 1: Proposed cyclization mechanism for kaurene synthase, a Class I DTS



Scheme 2: Biosynthetic Pathways from Geranylgeranyl Diphosphate (**1**) to *ent*-Kaurene (**3**), via *ent*-Copalyl Diphosphate (**2**) or to Abietadiene (**5**), via normal-Copalyl Diphosphate (**4**). In plants, the synthesis of kaurene requires two enzymes, *ent*-copalyl diphosphate synthase [*ent*-CPS] (for the conversion of **1** to **2**) and kaurene synthase [KS] (for the conversion of **2** to **3**). In grand fir, the conversion of **1** to **5** is accomplished by a single enzyme, abietadiene synthase [AS].



Scheme 3: Known labdane-related diterpenoid phytoalexins produced by rice (*O. sativa*).



Scheme 4: Known cyclization steps in labdane-related diterpenoid biosynthesis in rice. Reactions catalyzed by class II (tpsII) or class I (tpsI) terpene synthases are indicated, along with the corresponding classes of natural products derived from each of the named polycyclic hydrocarbon structures (dashed arrows indicate multiple enzymatic steps).

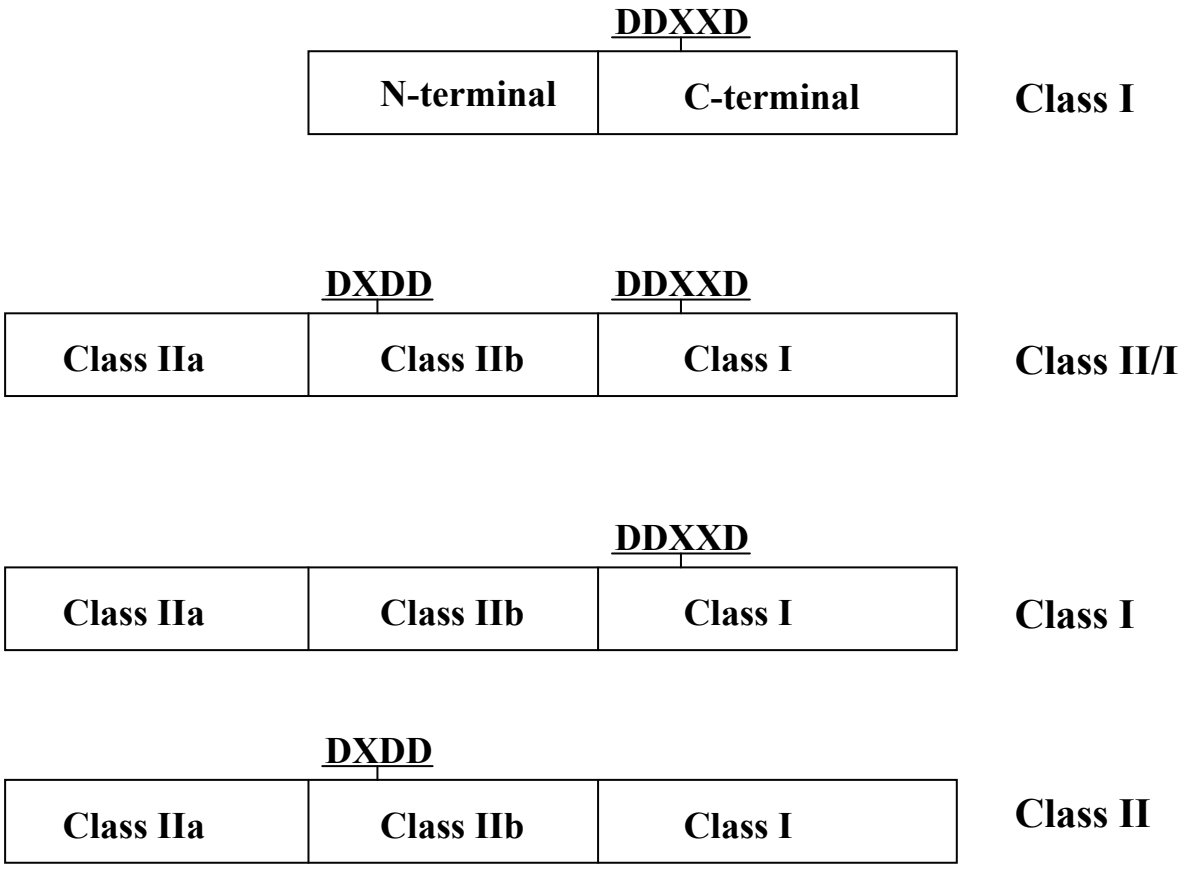


Figure 1: Domain structure of terpene synthases

Chapter II. Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis.

P. Ross Wilderman, Meimei Xu, Yinghua Jin, Robert M. Coates, Reuben J. Peters

ABSTRACT

Rice (*Oryza sativa*) produces momilactone diterpenoids as both phytoalexins and allelochemicals. Accordingly, the committed step in biosynthesis of these natural products is catalyzed by the class I terpene synthase that converts *syn*-copalyl diphosphate to the corresponding polycyclic hydrocarbon intermediate *syn*-pimara-7,15-diene. Here, a functional genomics approach was utilized to identify a *syn*-copalyl diphosphate specific 9ß-pimara-7,15-diene synthase (OsDTS2). This is the first identified terpene synthase with this particular substrate stereoselectivity and, by comparison with the previously described and closely related *ent*-copalyl diphosphate specific cassa-12,15-diene synthase (OsDTC1), provides a model system for investigating the enzymatic determinants underlying the observed difference in substrate specificity. Further, *OsDTS2* mRNA in leaves is upregulated by conditions that stimulate phytoalexin biosynthesis, but is constitutively expressed in roots, where momilactones are constantly synthesized as allelochemicals. Therefore, transcription of *OsDTS2* seems to be an important regulatory point for controlling production of these defensive compounds. Finally, the gene identified here as *OsDTS2* has previously been mapped at 14.3 cM on chromosome 4. The class II terpene synthase

producing *syn*-copalyl diphosphate from the universal diterpenoid precursor geranylgeranyl diphosphate was also mapped to this same region. These genes catalyze sequential cyclization steps in momilactone biosynthesis and seem to have been evolutionarily coupled by physical linkage and resulting co-segregation. Further, the observed correlation between physical proximity and common metabolic function indicates that other such class I and class II terpene synthase gene clusters may similarly catalyze consecutive reactions in shared biosynthetic pathways.

INTRODUCTION

Plants produce a vast and diverse array of low-molecular weight organic compounds, the overwhelming majority of which are secondary metabolites with non-essential, yet important functions such as defense (Croteau et al., 2000). For example, phytoalexins are produced in response to microbial infections and exhibit antimicrobial activity (VanEtten et al., 1994), while allelochemicals are secreted to the rhizosphere and suppress the growth of neighboring plants (Bais et al., 2004). Often found serving in such roles are terpenoids, which are particularly abundant in plant metabolism and form the largest class of natural products, exhibiting wide diversity in chemical structure and biological function (Croteau et al., 2000). Much of the structural variation within this class arises from the diverse carbon backbones formed by terpene synthases (cyclases). These divalent metal dependent enzymes carry out complex electrophilic cyclizations/rearrangements to create these diverse skeletal structures from relatively simple acyclic precursors (Davis and Croteau, 2000). Notably, production of a specific backbone structure either dictates, or at least severely restricts, the metabolic fate of that particular molecule. Thus, terpenoid natural products biosynthesis is often controlled by regulating terpene synthase activity [e.g., gibberellin biosynthesis; (Silverstone et al., 1997)].

A substantial fraction of the known terpenoids can be classified as labdane-related diterpenoids, a large group of over 5,000 known compounds that minimally contain the bicyclic hydrocarbon structure found in the labdane family of diterpenes. This structural core is formed by class II terpene synthases, which selectively produce specific stereoisomers of labdadienyl/copalyl diphosphate (CPP)[1] from the universal diterpenoid precursor geranylgeranyl diphosphate (GGPP). In addition, this core structure can be further modified and/or rearranged by stereoselective CPP specific class I terpene synthases, as in the related/derived structural families (e.g., gibberellins, abietanes, and (iso)pimaradienes). Thus, class II and class I terpene synthases act sequentially in catalyzing stereochemically coupled cyclization reactions to form labdane-related skeletal backbones (e.g., Figure 1).

While the two classes of terpene synthases are clearly phylogenetically related (Bohlmann et al., 1998a) and both catalyze electrophilic cyclization/rearrangement reactions, each utilizes completely distinct initiation mechanisms. Most commonly, terpene synthases (i.e., the class I enzymes) initiate their reactions through *ionization* of the allylic diphosphate moiety of their substrate. This divalent metal ion dependent reaction requires a DDXXD metal binding sequence that is the signature motif of class I terpene synthases (Davis and Croteau, 2000). In contrast, the class II enzymes initiate cyclization via *protonation* of the terminal carbon-carbon double bond in GGPP and contain a separately placed DXDD sequence rather than the class I associated DDXXD motif (Sun and Kamiya, 1994). Notably, prototypical class I enzymes are similar in size and contain two structurally defined domains (Starks et al., 1997; Whittington et al., 2002). However, some terpene synthases, in particular, all of those involved in labdane-related diterpenoid biosynthesis (Figure 2),

contain a large amount of additional N-terminal sequence termed the ‘insertional’ element [~240 residues; (Peters and Croteau, 2002)].

Rice provides a model system to investigate labdane-related diterpenoid biosynthesis, as this genomically characterized plant (Goff et al., 2002; Yu et al., 2002) produces a number of such natural products beyond the ubiquitous gibberellin growth hormones (Figure 1). These compounds include momilactones A & B (Kato et al., 1973; Cartwright et al., 1981), oryzalexins A-F (Akatsuka et al., 1985; Sekido et al., 1986; Kato et al., 1993, 1994), oryzalexin S (Kodama et al., 1992), and phytocassanes A-E (Koga et al., 1995; Koga et al., 1997). All of these natural products exhibit antimicrobial properties and are synthesized as part of the defensive response to leaf infection with the blast pathogenic fungus *Magneportha grisea*, thus, qualifying as phytoalexins (VanEtten et al., 1994). In addition, momilactones A & B also act as allelochemicals, as these were originally identified as dormancy factors from rice seed husks (Kato et al., 1973). More recently, momilactone B has been shown to be constitutively secreted from the roots of rice seedlings (Kato-Noguchi and Ino, 2003) as an allelopathic agent (Kato-Noguchi et al., 2002). Further, secretion of antimicrobial agents to the rhizosphere may also provide a competitive advantage for root establishment through local suppression of soil microorganisms (Bais et al., 2004).

Conveniently, rice leaves produce all of these secondary metabolites after UV-irradiation as well as blast fungal infection (Cartwright et al., 1977), providing a by-now standard method for inducing biosynthesis of these natural products (Kodama et al., 1988). For example, UV-irradiation induces biosynthesis of *ent*-sandaracopimaradiene, *syn*-pimara-7,15-diene, and *syn*-stemar-13-ene, the putative precursors to oryzalexins A-F, momilactones A & B, and oryzalexin S, respectively (Wickham and West, 1992). These polycyclic

diterpene hydrocarbons have been shown to be selectively produced via CPP of the corresponding stereochemistry [i.e., *ent* or *syn*; (Mohan et al., 1996)]. Recent work has identified the class I diterpene cyclase (OsDTC1) producing *ent*-cassa-12,15-diene, precursor to phytocassanes A-E (Yajima et al., 2004), stereoselectively from *ent*-CPP (Cho et al., 2004). In addition, a very recent report detailed the identification of rice gibberellin specific enzymatic genes by mutant plant phenotype and also mapped all the potentially relevant genes on the genome, revealing clustering of certain class II and class I terpene synthases (Sakamoto et al., 2004).[2] However, the genes involved in gibberellin biosynthesis are not clustered together. Because gene isolation and biochemical characterization were not otherwise reported, it is not known if the clustered genes operate sequentially in common metabolic pathways. We have functionally identified the three class II terpene synthases from rice; a *syn*-CPP synthase ($OsCPS_{syn}$) involved in phytoalexin/allelochemical biosynthesis (Xu et al., 2004), and two disparate *ent*-CPP synthases that are separately responsible for gibberellin ($OsCPS_{ent1}$) or phytoalexin ($OsCPS_{ent2}$) biosynthesis (Prisic et al., 2004). Here we report functional identification of a rice class I terpene synthase gene as a *syn*-CPP specific 9ß-pimara-7,15-diene synthase (OsDTS2). Both *OsDTS2* and $OsCPS_{syn}$ map to 14.3 cM on chromosome 4. Thus, their shared metabolic function, catalysis of consecutive cyclization reactions to initiate biosynthesis of the phytoalexin/allelochemical momilactones A & B, seems to have led to evolutionary coupling through physical linkage and resulting co-segregation.

RESULTS

Isolation of a class I terpene synthase cDNA from UV-irradiated rice leaves

The class I terpene synthases involved in labdane-related diterpenoid biosynthesis discriminate between different stereoisomers of CPP and, thus, provide a model system for investigating the underlying active site steric constraints. In addition, terpene synthases are conserved by taxonomic origin rather than biochemical function (Bohlmann et al., 1998a). Therefore, we were interested in identifying a *syn*-CPP specific class I terpene synthase from rice to complement the previously identified *ent*-CPP specific cassa-12,15-diene synthase (Cho et al., 2004). Towards this end, an initial class I terpene synthase with potential involvement in labdane-related diterpenoid biosynthesis (i.e., DDXXD motif and 'insertional' element) was found among the genes predicted from the finished sequence of rice chromosome 4 [accession CAD39507; (Feng et al., 2002)]. The corresponding sequence was readily amplified from UV-irradiated rice leaves, cloned, and verified by complete sequencing of two independent isolates, demonstrating ~99% identity to the predicted sequence. Presumably the few observed differences are a function of intersubspecies variation between the ssp. *japonica* sequenced by Feng et al. (2002) and ssp. *indica* used here (our sequence has been deposited into the various nucleotide sequence databases as accession AY616862).

Sequence comparison suggests a role in labdane-related diterpenoid biosynthesis

The cloned open reading frame encodes a protein of 840 amino acids that contains the 'insertional' element associated with labdane-related diterpene synthases (Figure 2A). Also present is the catalytically requisite class I DDXXD motif, as found in the original gene

prediction (accession CAD39507). Notably, the currently predicted gene product does not include the exon containing this metal binding motif (accession CAD39717), indicating that caution must be taken when analyzing predicted genes (i.e., the current prediction might be mistaken as a non-functional 'pseudo-gene'). Significantly, the DXDD motif required for class II cyclization is not found. Alignment of the complete amino acid sequence with that of known 'insertional' element containing class I terpene synthases (e.g., Figure 2B) demonstrated only weak similarity (<27% identity) to those not involved in labdane-related diterpenoid biosynthesis (Dudareva et al., 1996; Wildung and Croteau, 1996; Bohlmann et al., 1998b). Slightly higher similarity (~30% identity) was found with the identified gymnosperm bifunctional class II/I diterpene synthases, which are involved in labdane-related biosynthesis (Stofer Vogel et al., 1996; Schepmann et al., 2001; Martin et al., 2004). Further, the sequence is moderately similar (39-42% identity) to the known *ent*-kaurene synthases involved in gibberellin biosynthesis (Yamaguchi et al., 1996; Yamaguchi et al., 1998; Richman et al., 1999). Finally, comparison with the recently identified *ent*-CPP specific cassa-12,15-diene synthase (Cho et al., 2004) revealed significant homology (~51% identity). These results suggest that the isolated sequence is also a labdane-related (i.e., CPP specific) diterpene synthase (OsDTS2).

Functional characterization of OsDTS2 as *syn*-CPP specific 9ß-pimara-7,15-diene synthase

OsDTS2 was expressed both alone and as a fusion protein to glutathione-S-transferase (GST-OsDTS2). Recombinant preparations were assayed with GGPP, *ent*-CPP or *syn*-CPP as substrate and enzymatic activity detected using gas chromatography-mass spectrometry (GC-MS) analysis of organic extracts of the assays. Only the GST-OsDTS2 fusion protein exhibited appreciable amounts of activity, indicating the transit peptide (required for plastidial import in planta) hinders folding of the full-length preprotein in the absence of the stabilizing effect provided by the fused GST structure, as found in previous studies (Williams et al., 1998; Peters et al., 2000). Further, enzymatic turnover was only observed with *syn*-CPP, as no products were detected from reactions with GGPP or *ent*-CPP. Therefore, OsDTS2 is stereoselective and represents the first identified class I terpene synthase specific for *syn*-CPP. Finally, comparison with the known synthetic standards (Mohan et al., 1996) unambiguously identified the enzymatic product as *syn*-pimara-7,15-diene (Figure 3).

Expression pattern of OsDTS2 mRNA

The production of 9ß-pimara-7,15-diene from *syn*-CPP is the committed step in momilactone biosynthesis (Figure 1). Therefore, regulation of the corresponding activity (i.e., OsDTS2) provides a logical point for controlling the production of these specific labdane-related diterpenoid natural products. Previous review of the relevant literature has been used to suggest that plant secondary metabolism is most often regulated at the level of transcription (Peters and Croteau, 2004). This has been demonstrated for the preceding enzyme $OsCPS_{syn}$, where the corresponding mRNA increases prior to *syn*-labdane-related diterpenoid phytochemical accumulation in UV-irradiated rice leaves (Xu et al., 2004).

Nevertheless, it seemed likely that OsDTS2 would also be regulated, as OsDTC1 is similarly controlled (Cho et al., 2004), in addition to the observed transcriptional control of the upstream enzyme OsCPS$_{ent2}$ (Prisic et al., 2004). Such a control mechanism was investigated through the use of UV-irradiation, which has long been appreciated to induce phytoalexin biosynthesis in rice (Cartwright et al., 1977). Further, quantitative analysis of phytochemical accumulation for the detached leaf UV-irradiation induction method used here has been previously reported (Kodama et al., 1988). Hence, the ability of UV-irradiation to induce expression of *OsDTS2* mRNA in rice leaves was characterized, demonstrating transcriptional upregulation prior to phytoalexin accumulation (Figure 4). To verify that such transcriptional control is part of the normal regulatory mechanism for phytoalexin biosynthesis, it was further demonstrated that *OsDTS2* mRNA levels are also increased by methyl jasmonate (MeJA), an important plant defense signaling molecule (Farmer and Ryan, 1990). MeJA has also been previously demonstrated to induce phytoalexin biosynthesis in rice cell culture (Nojiri et al., 1996), as well as transcription of the phytoalexin specific class II terpene synthases OsCPS$_{ent2}$ and OsCPS$_{syn}$ (Prisic et al., 2004; Xu et al., 2004). Finally, as expected from its requisite role in constant production of an allelochemical (Kato-Noguchi and Ino, 2003), *OsDTS2* mRNA seems to be constitutively present in roots, albeit at a low level, again correlated with the expression pattern for OsCPS$_{syn}$ (Xu et al., 2004). These results strongly indicate that, in addition to the previously observed transcriptional regulation of OsCPS$_{syn}$ (Xu et al., 2004), biosynthesis of the phytoalexin/allelochemical momilactones is also more specifically controlled by transcriptional regulation of *OsDTS2*.

DISCUSSION

A class I terpene synthase containing sequence characteristics associated with labdane-related diterpene biosynthesis (*OsDTS2*) was found among the genes initially predicted from the finished sequence of rice chromosome 4 (Feng et al., 2002). The corresponding sequence was readily cloned from UV-irradiated rice leaves, which are known to produce a number of labdane-related diterpenoid phytoalexins (Kodama et al., 1988). Functional characterization demonstrated that OsDTS2 is specific for *syn*-CPP, producing 9ß-pimara-7,15-diene (Figure 3). Notably, OsDTS2 represents the first identified terpene synthase to exhibit stereoselectivity for *syn*-, rather than *ent*- or normal CPP. In addition, OsDTS2 is ~51% identical to the *ent*-CPP specific OsDTC1 identified by Cho et al. (2004). Hence, these provide an ideal model system for comparative investigation of the differential enzymatic determinants specifying the observed substrate selectivity for configurationally varied stereoisomers of CPP (Figure 5). Further, radio-tracer biosynthetic studies indicate that rice produces at least seven different polycyclic diterpenes derived from *ent*- or *syn*-CPP (Mohan et al., 1996), and the essentially single product output of OsDTC1 and OsDTS2 suggests that there will be individual stereospecific class I diterpene synthases responsible for each of these. From the extensive genomic and cDNA sequence information available for rice (Goff et al., 2002; Yu et al., 2002; Kikuchi et al., 2003) there are a total of nine putative 'insertional' element containing class I terpene synthases, as indicated by Cho et al. (2004) and Sakamoto et al. (2004), as well as our own searches of the relevant databases. Thus, functional characterization of these genes is expected to further increase the utility of this model system by providing additional sequences and enzymatic targets for comparative analysis.

Notably, production of *syn*-pimara-7,15-diene is the committed step in momilactone biosynthesis, the end products of which exhibit both phytoalexin and allelochemical properties (Cartwright et al., 1981; Kato-Noguchi and Ino, 2003). Therefore, OsDTS2 provides a logical regulatory target for controlling production of these important phytochemicals. Expression analysis demonstrated that *OsDTS2* mRNA is only found in association with momilactone biosynthesis (Figure 4). Accordingly, *OsDTS2* is constitutively expressed in the roots for constant allelochemical production, but is only found in leaves under conditions that induce phytoalexin biosynthesis (i.e., UV-irradiation or exposure to methyl jasmonate). These results strongly indicate that OsDTS2 activity is controlled by transcriptional regulation.

Intriguingly, in their investigation of rice gibberellin biosynthesis Sakamoto et al. (2004) mapped all terpene synthases with potential involvement in labdane-related diterpenoid biosynthesis to their chromosomal locations, demonstrating co-segregation and relatively close physical proximity for a number of these genes. In particular, the previously identified *syn*-CPP producing $OsCPS_{syn}$ (Xu et al., 2004) and the *syn*-CPP substrate specific *OsDTS2* identified here have both been mapped at 14.3 cM on chromosome 4 and are within 120 kb of each other. These enzymes catalyze sequential cyclization reactions to initiate momilactone biosynthesis. Therefore, these two physically linked genes act in a common metabolic pathway, similar to the grouping of presumably consecutively acting prenyltransferases and terpene synthases noted in the genome of *Arabidopsis thaliana* (Aubourg et al., 2002). Co-segregation of these stereochemically coupled class II and class I terpene synthases provides an evolutionary mechanism linking their shared biosynthetic functions. Such correlation between physical proximity and sequential function in a common

biosynthetic pathway is also consistent with our identification of a specific role for the *ent*-CPP producing OsCPS$_{ent2}$ in secondary metabolism (Prisic et al., 2004), since both *OsCPS*$_{ent2}$ and the *ent*-CPP substrate specific *OsDTC1* identified by Cho et al. (2004) have been mapped at 86 cM on chromosome 2. Two other class I terpene synthases also map to this region, with all four genes being found within a stretch of 150 kb. Hence, we speculate that these two additional enzymes will also function in *ent*-labdane-related diterpenoid secondary metabolism (i.e., specifically utilize *ent*-CPP), forming another functional grouping linked by stereochemically coupled sequential cyclization reactions. Finally, both of these terpene synthase gene clusters contain retrotransposable elements, suggesting a means by which duplication and, hence, diversification of these genes and/or clusters may have occurred. Thus, it appears the correlation between physical proximity and shared metabolic function of terpene synthases has important implications for the evolutionary mechanism underlying development of the extensive and diverse labdane-related diterpenoid secondary metabolism exhibited by rice.

CONCLUSION

In summary, we have used a functional genomics approach to identify a 9ß-pimara-7,15-diene synthase (OsDTS2). Notably, the observed specificity for *syn*-CPP represents a novel substrate stereoselectivity. Thus, OsDTS2, in combination with the previously identified and relatively closely related (~51% identity) *ent*-CPP specific OsDTC1 (Cho et al., 2004), offers an ideal model system for future structure/function investigations of steric

constraints within the active site of class I terpene synthases. In addition, OsDTS2 catalyzes the committed step in biosynthesis of the phytoalexin/allelochemical momilactones, and its transcription seems to represent an important control point for regulation of these important metabolic processes. Finally, functional characterization of *OsDTS2*, along with the previous identification of $OsCPS_{syn}$ (Xu et al., 2004), demonstrates that the previously reported physical proximity and co-segregation of these genes reflects their consecutive action in a common metabolic pathway (i.e., momilactone biosynthesis). We further speculate that other such terpene synthase gene clusters may also share metabolic function in catalyzing sequential, stereochemically linked cyclization reactions.

MATERIALS AND METHODS

Chemicals. Synthesis of (*E*,*E*,*E*)-geranylgeranyl diphosphate (GGPP), *ent*- and *syn*-copalyl diphosphate (CPP), and the polycyclic hydrocarbon standards *ent*-kaurene, *ent*-sandaracopimaradiene, *syn*-pimara-7,15-diene, and *syn*-stemar-13-ene have been previously described (Mohan et al., 1996). Unless otherwise noted, all other chemicals were purchased from Fisher Scientific.

Plant material. Rice plants (*Oryza sativa* L. ssp. *indica* cv. IR24) and seedlings (ssp. *japonica* cv. Nipponbare) were those previously described (Xu et al., 2004). Briefly, detached leaves from four-week old greenhouse grown plants were UV-irradiated at 254 nm from 15 cm distance for 25 min. and then incubated for the indicated period of time under dark humid conditions at 30°C. Seedlings were germinated from surface sterilized seeds under sterile, humid conditions at 30°C in the dark for a week. The seedlings then underwent MeJA treatment, being sprayed with ~2 mL 0.1% Tween 20 ± 0.5 mM MeJA per gram of

plant weight, and the seedlings then incubated for two more days under the same conditions. RNA was isolated using Concert Plant Reagent (Invitrogen; Carlsbad, CA, USA). Semi-quantitative RT-PCR expression analysis, using 0.5 µg total RNA and *OsDTS2*-specific primers or QuantumRNA 18S standard primers (Ambion; Austin, TX, USA), was also carried out as described by Xu et al. (2004).

Cloning. A putative class I terpene synthase involved in labdane-related biosynthesis was identified by a BLAST search of the GenBank database (www.ncbi.nlm.nih.gov:80/BLAST/) with the amino acid sequence of *ent*-kaurene synthase from *Arabidopsis* (Yamaguchi et al., 1998). RT-PCR reactions were performed to verify expression of the predicted gene in UV-irradiated leaves by generating a fragment of the corresponding sequence. This was verified by cloning into pCR-Zero-Blunt (Invitrogen) and complete sequencing. These primers were also then used for the semi-quantitative RT-PCR expression analysis. The complete open reading frame for OsDTS2 was then amplified from total RNA in an RT-PCR reaction using the GeneRacer kit (Invitrogen), cloned into pENTR/SD/D-TOPO (Invitrogen), and verified by complete sequencing. OsDTS2 was then transferred by directional recombination to the T7 based expression vectors pDEST14 and pDEST15 (Gateway system, Invitrogen).

Recombinant expression and functional characterization. Expression was carried out with the BL21-derived C41 strain (Miroux and Walker, 1996), as described for $OsCPS_{syn}$ by Xu et al. (2004). Briefly, cells were grown to mid-log phase at 37°C then shifted to 16°C for 1-2 h prior to induction (1 mM IPTG) and overnight expression. The cells were harvested by centrifugation (3000*g*, 20 min, 5°C), re-suspended in 1 mL of lysis buffer (50 mM Bis-Tris, pH 6.8, 150 mM KCl, 10 mM $MgCl_2$, 1 mM DTT), and lysed by mild sonication on ice

(Branson sonifier 450: 20 s, continuous, output setting 5). The resulting lysates were cleared by centrifugation (15,000*g*, 30 min, 5°C) and filtration (0.8 μ) to yield recombinant soluble extracts. Enzymatic assays were performed with these preparations under standard conditions defined for diterpene synthase activity [e.g., (Peters et al., 2000)]. Briefly, reactions with ~50 μM substrate (GGPP, *ent*-, or *syn*-CPP) were carried out in assay buffer (50 mM HEPES, pH 7.2, 100 mM KCl, 7.5 mM $MgCl_2$, 5% glycerol, and 5 mM DTT) with 25 μL of recombinant protein in a total volume of 0.2 mL. The reaction was allowed to proceed for 3 h at 30°C prior to extraction with hexanes. GC-MS analysis was performed using an HP-5 column on an Agilent (Palo Alto, CA, USA) 6890N GC instrument with 5973N mass selective detector. Samples (5 μL) were injected at 40°C in the splitless mode. After holding 3 min. at 40°C, the temperature was increased at 20°C/min. to 300°C, where it was held for 3 min. MS data was collected from 50 to 500 *m/z* during the temperature ramp. The retention time (RT) and MS pattern were compared to those for *syn*-stemar-13-ene and *syn*-pimara-7,15-diene, as well as sandaracopimaradiene and *ent*-kaurene.

Sequence analysis and alignments. Sequence alignments and identity calculations were performed with the AlignX program in the Vector NTI software package (Invitrogen), using standard parameters. OsDTS2 was the reference sequence in all cases. The class I terpene synthases not involved in labdane-related diterpenoid biosynthesis, yet containing 'insertional' elements, are linalool synthase from *Clarkia breweri* (Dudareva et al., 1996), taxadiene synthase from *Taxus brevifolia* (Wildung and Croteau, 1996), and bisabolene synthase from *Abies grandis* (Bohlmann et al., 1998b). Bifunctional class II/I terpene synthases producing labdane-related diterpenes of normal stereochemistry are those from the gymnosperms *Abies grandis* (Stofer Vogel et al., 1996), *Ginkgo biloba* (Schepmann et al.,

2001), and *Picea abies* (Martin et al., 2004). The *ent*-kaurene synthases were those from *Curcurbita maxima* (Yamaguchi et al., 1996), *Arabidopsis thaliana* (Yamaguchi et al., 1998), and *Stevia rebaudiana* (Richman et al., 1999). Other rice class I terpene synthases with potential involvement in labdane-related diterpenoid biosynthesis were identified by BLAST queries of GenBank, and the cDNA databases at KOME (cdna01.dna.affrc.go.jp/cDNA/) and TIGR (tigrblast.tigr.org/tgi/), using *Arabidopsis ent*-kaurene synthase as the probe sequence.

ACKNOWLEDGEMENTS

The authors thank Dr. Adam Bogdanove and his laboratory for supplying rice plants and seeds, Dr. Robert Thornburg for methyl jasmonate, and Dana J. Morrone for critically reading the manuscript.

REFERENCES

Akatsuka T, Kodama O, Sekido H, Kono Y, Takeuchi S (1985) Novel Phytoalexins (Oryzalexins A, B, and C) Isolated from Rice Blast Leaves Infected with *Pyricularia oryzae*. Part I: Isolation, Characterization and Biological Activities of Oryzalexins. Agric. Biol. Chem. **49:** 1689-1694

Aubourg S, Lecharny A, Bohlmann J (2002) Genomic analysis of terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. Mol. Genet. Genomics **267:** 730-745

Bais HP, Park S-W, Weir TL, Callaway RM, Vivanco JM (2004) How plants communicate using the underground information superhighway. Trends Plant Sci. **9:** 26-32

Bohlmann J, Crock J, Jetter R, Croteau RB (1998b) Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (*E*)-alpha-bisabolene synthase from grand fir (*Abies grandis*). Proc Natl Acad Sci U S A. **95:** 6756-6761

Bohlmann J, Meyer-Gauen G, Croteau R (1998a) Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc. Natl. Acad. Sci. USA **95:** 4126-4133

Cartwright DW, Langcake P, Pryce RJ, Leworthy DP, Ride JP (1977) Chemical activation of host defence mechanisms as a basis for crop protection. Nature **267:** 511-513

Cartwright DW, Langcake P, Pryce RJ, Leworthy DP, Ride JP (1981) Isolation and characterization of two phytoalexins from rice as momilactones A and B. Phytochemistry **20:** 535-537

Cho E-M, Okada A, Kenmoku H, Otomo K, Toyomasu T, Mitsuhashi W, Sassa T, Yajima A, Yabuta G, Mori K, Oikawa H, Toshima H, Shibuya N, Nojiri H, Omori T, Nishiyama M, Yamane H (2004) Molecular cloning and characterization of a cDNA encoding *ent*-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor. Plant J. **37:** 1-8

Croteau R, Kutchan TM, Lewis NG (2000) Natural Products (Secondary Metabolites). *In* B Buchanan, W Gruissem, R Jones, eds, Biochemistry & Molecular Biology of Plants. Am. Soc. Plant Biologists, Rockville, MD, USA, pp 1250-1318

Davis EM, Croteau R (2000) Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes. Top. Curr. Chem. **209:** 53-95

Dudareva N, Cseke L, Blanc VM, Pichersky E (1996) Evolution of Floral Scent in *Clarkia*: Novel Patterns of *S*-Linalool Synthase Gene Expression in the *C. breweri* Flower. Plant Cell **8:** 1137-1148

Farmer EE, Ryan CA (1990) Interplant communication: Airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. Proc Natl Acad Sci U S A. **87:** 7713-7716

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X, Jia P, Zhang Y, Zhao Q, Ying K, Yu S, Tang Y, Weng Q, Zhang L, Lu Y, Mu J, Lu Y, Zhang LS, Yu Z, Fan D, Liu X, Lu T, Li C, Wu Y, Sun T, Lei H, Li T, Hu H, Guan J, Wu M, Zhang R, Zhou B, Chen Z, Chen L, Jin Z, Wang R, Yin H, Cai Z, Ren S, Lv G, Gu W, Zhu G, Tu Y, Jia J, Zhang Y, Chen J, Kang H, Chen X, Shao C, Sun Y, Hu Q, Zhang X, Zhang W, Wang L, Ding C, Sheng H, Gu J, Chen S, Ni L, Zhu F, Chen W, Lan L, Lai Y, Cheng Z, Gu M, Jiang J, Li J, Hong G, Xue Y, Han B (2002) Sequence and analysis of rice chromosome 4. Nature **420:** 316-320

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun WL, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science **296:** 92-100

Kato H, Kodama O, Akatsuka T (1993) Oryzalexin E, a diterpene phytoalexin from UV-irradiated rice leaves. Phytochemistry **33:** 79-81

Kato H, Kodama O, Akatsuka T (1994) Oryzalexin F, a diterpene phytoalexin from UV-irradiated rice leaves. Phytochemistry **36:** 299-301

Kato T, Kabuto C, Sasaki N, Tsunagawa M, Aizawa H, Fujita K, Kato Y, Kitahara Y (1973) Momilactones, growth inhibitors from rice, *Oryza sativa* L. Tetrahedron Lett. **39:** 3861-3864

Kato-Noguchi H, Ino T (2003) Rice seedlings release momilactone B into the environment. Phytochemistry **63:** 551-554

Kato-Noguchi H, Ino T, Sata N, Yamamura S (2002) Isolation and identification of a potent allelopathic substance in rice root exudates. Physiologia Plantarum **115:** 401-405

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, Hotta I, Kojima K, Namiki T, Ohneda E, Yahagi W, Suzuki K, Li CJ, Ohtsuki K, Shishiki T, Otomo Y, Murakami K, Iida Y, Sugano S, Fujimura T, Suzuki Y, Tsunoda Y, Kurosaki T, Kodama T, Masuda H, Kobayashi M, Xie Q, Lu M, Narikawa R, Sugiyama A, Mizuno K, Yokomizo S, Niikura J, Ikeda R, Ishibiki J, Kawamata M, Yoshimura A, Miura J, Kusumegi T, Oka M, Ryu R, Ueda M, Matsubara K, Kawai J, Carninci P, Adachi J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashidume W, Hayatsu N, Imotani K, Ishii Y, Itoh M, Kagawa I, Kondo S, Konno H, Miyazaki A, Osato N, Ota Y, Saito R, Sasaki D, Sato K, Shibata K, Shinagawa A, Shiraki T, Yoshino M, Hayashizaki Y (2003) Collection, Mapping, and Annotation of Over 28,000 cDNA Clones from *japonica* Rice. Science **301:** 376-379

Kodama O, Li WX, Tamogami S, Akatsuka T (1992) Oryzalexin-S, a novel stemarane-type diterpene rice phytoalexin. Biosci. Biotechnol. Biochem. **56:** 1002-1003

Kodama O, Suzuki T, Miyakawa J, Akatsuka T (1988) Ultraviolet-induced accumulation of phytoalexins in rice leaves. Agric. Biol. Chem. **52:** 2469-2473

Koga J, Ogawa N, Yamauchi T, Kikuchi N, Ogasawara N, Shimura M (1997) Functional moiety for the antifungal activity of phytocassane E, a diterpene phytoalexin from rice. Phytochemistry **44:** 249-253

Koga J, Shimura M, Oshima K, Ogawa N, Yamauchi T, Ogasawara N (1995) Phytocassanes A, B, C, and D, Novel Diterpene Phytoalexins from Rice, *Oryza sativa* L. Tetrahedron **51:** 7907-7918

Martin DM, Faldt J, Bohlmann J (2004) Functional Characterization of Nine Norway Spruce *TPS* genes and Evolution of Gymnosperm Terpene Synthases of the *TPS-d* Subfamily. Plant Physiol. **in press**

Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. Mol. Biol. **260:** 289-298

Mohan RS, Yee NKN, Coates RM, Ren YY, Stamenkovic P, Mendez I, West CA (1996) Biosynthesis of cyclic diterpene hydrocarbons in rice cell suspensions: conversion of 9,10-*syn*-labda-8(17),13-dienyl diphosphate to 9ß-pimara-7,15-diene and stemar-13-ene. Arch. Biochem. Biophys. **330:** 33-47

Nojiri H, Sugimora M, Yamane H, Nishimura Y, Yamada A, Shibuya N, Kodama O, Murofushi N, Omori T (1996) Involvement of Jasmonic Acid in Elicitor-Induced Phytoalexin Production in Suspension-Cultured Rice Cells. Plant Physiol. **110:** 387-392

Peters RJ, Carter OA, Zhang Y, Matthews BW, Croteau RB (2003) Bifunctional Abietadiene Synthase: Mutual Structural Dependence of the Active Sites for Protonation-Initiated and Ionization-Initiated Cyclizations. Biochemistry **42:** 2700-2707

Peters RJ, Croteau RB (2002) Abietadiene Synthase Catalysis: Conserved Residues Involved in Protonation-Initated Cyclization of Geranylgeranyl Diphosphate to (+)-Copalyl Diphosphate. Biochemistry **41:** 1836-1842

Peters RJ, Croteau RB (2004) Metabolic Engineering of Plant Secondary Metabolism. *In* G Kishore, ed, Handbook of Plant Biotechnology: Applications of Plant Biotechnology in Agriculture, the Pharmaceutical Industry, and Other Industries, Vol 2. John Wiley & Sons Ltd, London, pp 609-628

Peters RJ, Flory JE, Jetter R, Ravn MM, Lee H-J, Coates RM, Croteau RB (2000) Abietadiene Synthase from Grand Fir (*Abies grandis*): Characterization and Mechanism of Action of the "Pseudomature" Recombinant Enzyme. Biochemistry **39:** 15592-15602

Prisic S, Xu M, Wilderman PR, Peters RJ (2004) Disparate *ent*-copalyl diphosphate synthases from rice suggest an early split between gibberellin and related secondary metabolism in the cereal family. Plant J. **136:** 4228-4236

Richman AS, Gijzen M, Starratt AN, Yang Z, Brandle JE (1999) Diterpene synthesis in *Stevia rebaudiana*: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway. Plant J. **19:** 411-421

Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K, Miyao A, Hirochika H, Kitano H, Ashikari M, Matsuoka M (2004) An Overview of Gibberellin Metabolism Enzyme Genes and Their Related Mutants in Rice. Plant Physiol. **134:** 1642-1653

Schepmann HG, Pang J, Matsuda SP (2001) Cloning and Characterization of *Ginkgo biloba* Levopimaradiene Synthase, Which Catalyzes the First Committed Step in Ginkolide Biosynthesis. Arch. Biochem. Biophys. **392:** 263-269

Sekido H, Endo T, Suga R, Kodama O, Akatsuka T, Kono Y, Takeuchi S (1986) Oryzalexin D (3,7-Dihydroxy-(+)-sandaracopimaradiene), a New Phytoalexin Isolated from Blast-infected Rice Leaves. J. Pesticide Sci. **11:** 369-372

Silverstone AL, Chang C-W, Krol E, Sun T-P (1997) Developmental regulation of the gibberellin biosynthetic gene GA1 in *Arabidopsis thaliana*. Plant J. **12:** 9-19

Starks CM, Back K, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science **277:** 1815-1820

Stofer Vogel B, Wildung MR, Vogel G, Croteau R (1996) Abietadiene Synthase from Grand Fir (*Abies grandis*). J. Biol. Chem. **271:** 23262-23268

Sun T-P, Kamiya Y (1994) The Arabidopsis GA1 locus encodes the cyclase *ent*-kaurene synthetase A of gibberellin biosynthesis. Plant Cell **6:** 1509-1518

VanEtten HD, Mansfield JW, Bailey JA, Farmer EE (1994) Two classes of plant antibiotics: phytoalexins versus 'phytoanticipins'. Plant Cell **6:** 1191-1192

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci U S A. **99:** 15375-15380

Wickham KA, West CA (1992) Biosynthesis of Rice Phytoalexins: Identification of Putative Diterpene Hydrocarbon Precursors. Arch. Biochem. Biophys. **293:** 320-332

Wildung MR, Croteau RB (1996) A cDNA clone for taxadiene synthase, the diterpene cyclase that catalyzes the committed step of taxol biosynthesis. J. Biol. Chem. **271:** 9201-9204

Williams DC, McGarvey DJ, Katahira EJ, Croteau R (1998) Truncation of Limonene Synthase Preprotein Provides a Fully Active 'Pseudomature' Form of This Monoterpene Cyclase and Reveals the Function of the Amino-Terminal Arginine Pair. Biochemistry **37:** 12213-12220

Xu M, Hillwig ML, Prisic S, Coates RM, Peters RJ (2004) Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. Plant J. **39:** 309-318

Yajima A, Mori K, Yabuta G (2004) Total synthesis of *ent*-cassa-12,15-diene, a putative precursor of rice phytoalexins, phytocassanes A-E. Tetrahedron Lett. **45:** 167-169

Yamaguchi S, Saito T, Abe H, Yamane H, Murofushi N, Kamiya Y (1996) Molecular cloning and characterization of a cDNA encoding the gibberellin biosynthetic enzyme *ent*-kaurene synthase B from pumpkin (*Cucurbita maxima* L.). Plant J **10:** 101-111

Yamaguchi S, Sun T, Kawaide H, Kamiya Y (1998) The GA2 Locus of *Arabidopsis thaliana* Encodes *ent*-Kaurene Synthase of Gibberellin Biosynthesis. Plant Physiol. **116:** 1271-1278

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79-92

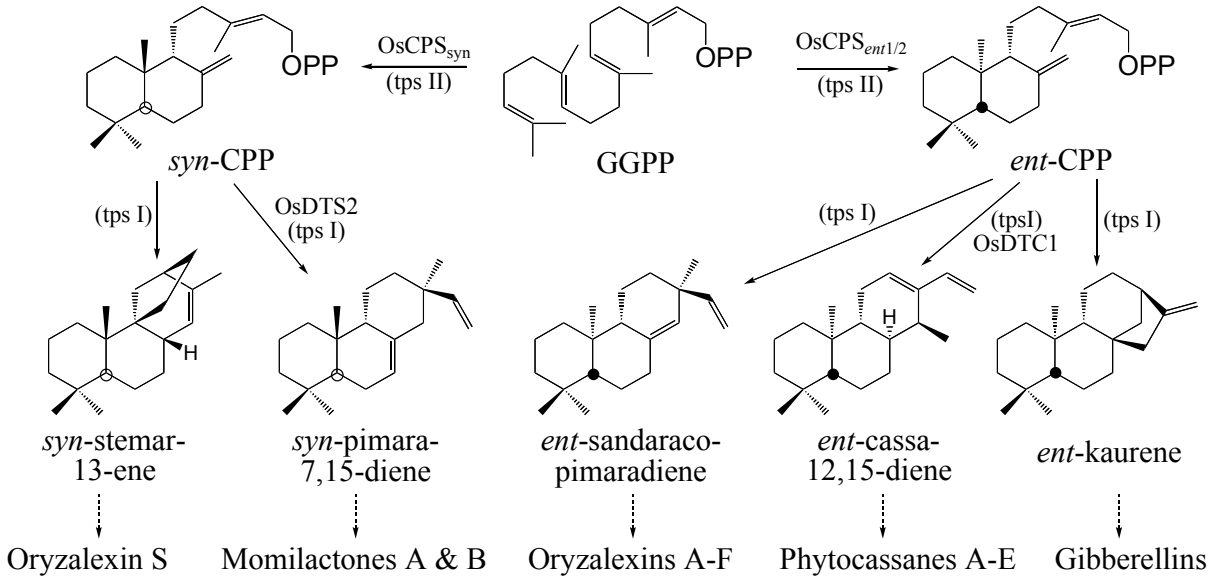
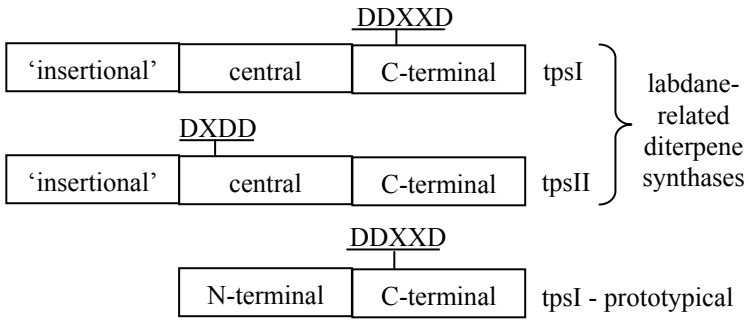


Figure 1: Known cyclization steps in labdane-related diterpenoid biosynthesis in rice. Reactions catalyzed by class II (tpsII) or class I (tpsI) terpene synthases are indicated, along with the corresponding classes of natural products derived from each of the named polycyclic hydrocarbon structures (dashed arrows indicate multiple enzymatic steps).

A)


DDXXD
'insertional'
central
C-terminal
tpsI
labdane-related diterpene synthases
DXDD
'insertional'
central
C-terminal
tpsII
DDXXD
N-terminal
C-terminal
tpsI - prototypical

B)

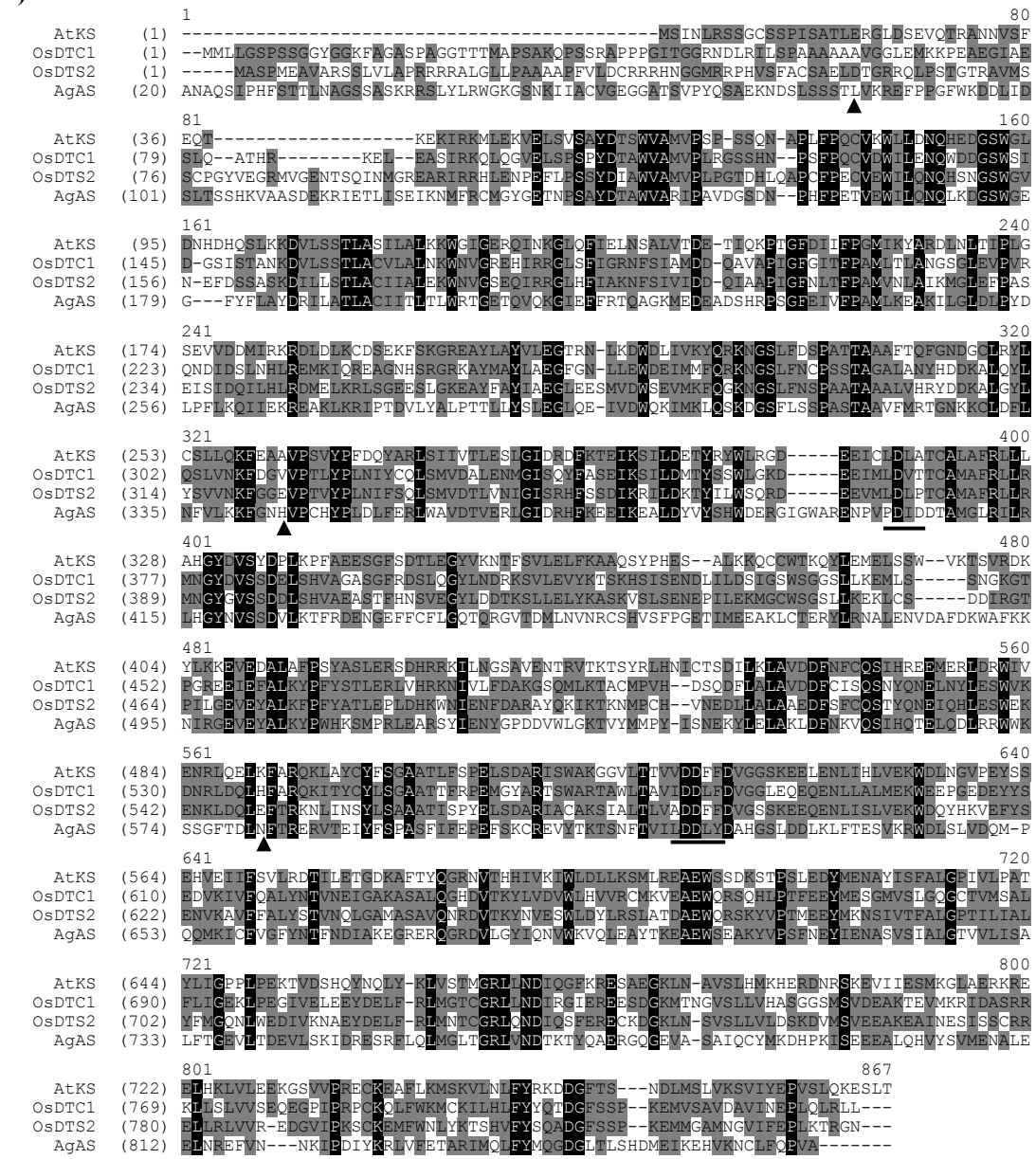

AtKS
OsDTC1
OsDTS2
AgAS

Figure 2: Terpene synthase comparisons. A) Domain schematic for prototypical class I terpene synthases (tpsI) and the class I and class II (tpsII) enzymes involved in labdane-related diterpene biosynthesis. Modeled on the structures determined for the typical class I *epi*-aristolochene and bornyl diphosphate synthases (Starks et al., 1997; Whittington et al., 2002), along with the additional 'insertional' sequence element. Approximate locations of the aspartate-rich motifs are also indicated. For clarity, the transit sequence required in mono- and di-terpene synthases for their plastidal location in planta is not shown. B) Sequence comparison of OsDTS2 with other class I terpene synthases. Alignment with the previously identified rice *ent*-cassa-12,15-diene synthase (OsDTC1), *Arabidopsis thaliana ent*-kaurene synthase (AtKS), and bifunctional class II/I cyclase from *Abies grandis* abietadiene synthase (AgAS). The approximate domain boundaries, based on previous analysis of AgAS (Peters et al., 2003), are marked with arrowheads and the class II associated DXDD and class I associated DDXXD motifs are underlined.

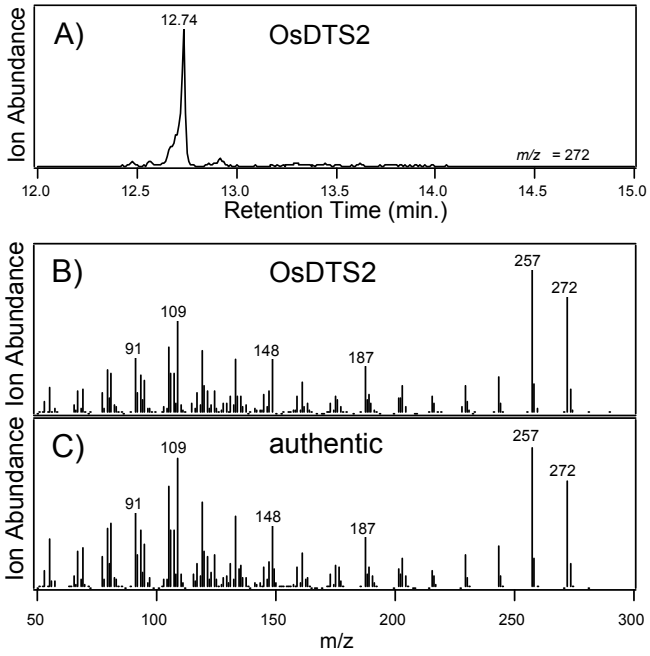


Figure 3: Production of *syn*-pimara-7,15-diene. A) GC-MS analysis (272 *m/z* extracted ion chromatograph) of the product derived from *syn*-CPP by OsDTS2. B) Mass spectrum of the GC-MS 272 *m/z* chromatograph peak for OsDTS2 (RT = 12.74 min.). C) Mass spectrum from an authentic standard for 9ß-pimara-7,15-diene, which also exhibits RT = 12.74 min.

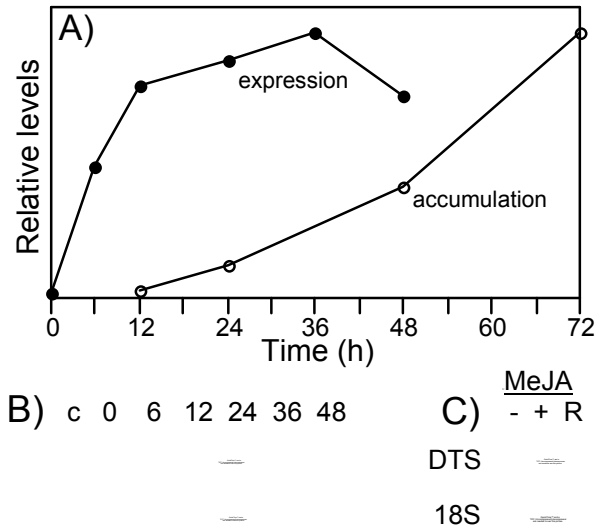


Figure 4: OsDTS2 expression analysis. A) Graphical comparison of OsDTS2 mRNA levels (closed circles) and momilactone accumulation [open circles; as described by (Kodama et al., 1988)], in UV-irradiated detached leaves. Semi-quantitative RT-PCR analysis of OsDTS2 mRNA expression levels is shown in B) and C). Specific bands corresponding to the 18S rRNA control and OsDTS2 (DTS) are indicated. B) Expression in response to UV-irradiation. Time (hours) after exposure is indicated (c = control leaves after ~18 hours). C) Expression in untreated four-week old plant roots (R), or in germinated seedlings in response to application of 0.5 mM methyl jasmonate (+MeJA) or water control (-MeJA).

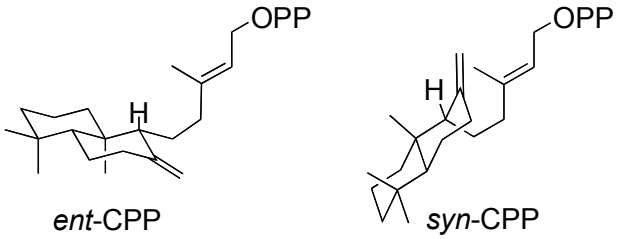


Figure 5: Configurational differences between *ent*- and *syn*-CPP. Both are depicted in the sterically less hindered boat-boat conformation, with similar positions for the diphosphate group whose relative position in the active site is fixed by the conserved divalent metal ion binding sites. OsDTC1 and OsDTS2 are able to distinguish between these two stereoisomers based on these configurational differences.

Chapter III. Functional Characterization of the Rice Kaurene Synthase-Like Gene Family

Reproduced with permission from **Xu, M., Wilderman, P. R., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N. M., Coates, R. M., Peters, R. J.**, 2007. Functional Characterization of the Rice Kaurene Synthase-Like Gene Family. *Phytochemistry* 68, 312-326.

Meimei Xu, P. Ross Wilderman, Dana Morrone, Jianjun Xu, Arnab Roy, Marcia Margis-Pinheiro, Narayana M. Upadhyaya, Robert M. Coates, Reuben J. Peters

Abstract

The rice (*Oryza sativa*) genome contains a family of kaurene synthase-like genes (*OsKSL*) presumably involved in diterpenoid biosynthesis. While a number of OsKSL enzymes have been functionally characterized, several have not been previously investigated, and the gene family has not been broadly analyzed. Here we report cloning of several *OsKSL* genes and functional characterization of the encoded enzymes. In particular, we have verified the expected production of *ent*-kaur-16-ene by the gibberellin phytohormone biosynthesis associated OsKS1 and demonstrated that OsKSL3 is a pseudo-gene, while OsKSL5 and OsKSL6 produce *ent*-(iso)kaur-15-ene. Similar to previous reports, we found that our sub-species variant of OsKSL7 produces *ent*-cassa-12,15-diene, OsKSL10 produces *ent*-(sandaraco)pimar-8(14),15-diene, and OsKSL8 largely *syn*-stemar-13-ene, although we also identified *syn*-stemod-12-ene as an alternative product formed in ~20% of the reactions catalyzed by OsKSL8. Along with our previous reports identifying OsKSL4 as a *syn*-pimara-

7,15-diene synthase and OsKSL11 as a *syn*-stemod-13(17)-ene synthase, this essentially completes biochemical characterization of the *OsKSL* gene family, enabling broader analyses. For example, because several OsKSL enzymes are involved in phytoalexin biosynthesis and their gene transcription is inducible, promoter analysis was used to identify a pair of specifically conserved motifs that may be involved in transcriptional up-regulation during the rice plant defense response. Also examined is the continuing process of gene evolution in the *OsKSL* gene family, which is particularly interesting in the context of very recently reported data indicating that a *japonica* sub-species variant of OsKSL5 produces *ent*-pimara-8(14),15-diene, rather than the *ent*-(iso)kaur-15-ene produced by the *indica* sub-species variant analyzed here.

1. Introduction

Rice is an important food crop, and has become a model plant for the cereal plant family with the recent availability of draft genome sequences (Goff et al., 2002; Yu et al., 2002), as well as large numbers of defined full-length cDNAs (Kikuchi et al., 2003). This extensive sequence information has enabled functional genomics based approaches towards elucidating the biosynthetic machinery underlying rice metabolism. One specific area of interest is the production of natural products with antimicrobial activity, which are termed phytoalexins if their biosynthesis is induced by microbial infection and phytoanticipins if their biosynthesis is constitutive (VanEtten et al., 1994), as the production of these small organic compounds is associated with resistance to microbial diseases such as that caused by the agronomically devastating rice blast pathogen *Magneporthe grisea*.

Extensive phytochemical investigation has demonstrated that rice produces a number of phytoalexins in response to *M. grisea* infection (Peters, 2006). Intriguingly, the rice phytoalexins, with the exception of the flavonoid sakuranetin, all fall into the large family of labdane-related diterpenoid natural products characterized by minimally containing a labdane type bicyclic core structure. Thus, along with the ubiquitous gibberellin phytohormone, rice produces more than 10 other labdane-related diterpenoids as phytoalexins. These are momilactones A & B (Cartwright et al., 1977, 1981), oryzalexins A-F (Akatsuka et al., 1985; Kato et al., 1993, 1994; Sekido et al., 1986), oryzalexin S (Kodama et al., 1992), and phytocassanes A-E (Koga et al., 1997; Koga et al., 1995). In addition, momilactone B is constitutively secreted from rice roots and acts as an allelochemical in suppressing germination in nearby seeds (Kato-Noguchi and Ino, 2003; Kato-Noguchi et al., 2002). The identified rice labdane-related diterpenoid natural products fall into five structurally related groups (Fig. 1), with the gibberellins being elaborated from *ent*-kaurene, oryzalexins A-F from *ent*-sandaracopimaradiene, phytocassanes A-E from *ent*-cassadiene, oryzalexin S from *syn*-stemodene, and momilactones A and B from *syn*-pimaradiene (Mohan et al., 1996; Wickham and West, 1992; Yajima et al., 2004).

Labdane-related diterpenoids share an unusual biogenetic origin, as their biosynthesis is uniquely initiated by a consecutive pair of terpene synthase catalyzed reactions. In the first reaction, the characteristic bicyclic core structure is formed by class II labdane-related diterpene cyclases, which catalyze protonation-initiated cyclization of the universal diterpenoid precursor (*E,E,E*)-geranylgeranyl diphosphate (GGPP, **1**) to produce a specific stereoisomer of labdadienyl/copalyl diphosphate (e.g. **2** and **3**) or rearranged derivative structure such as clerodanyl diphosphate (MacMillan and Beale, 1999). The resulting

cyclized diphosphate compounds can then be further cyclized and/or rearranged by more typical class I terpene synthases, which initiate catalysis via ionization of the allylic pyrophosphate group (Davis and Croteau, 2000). Notably, class I labdane-related diterpene synthases exhibit stereospecificity, e.g. all of the identified copalyl diphosphate (CPP) specific enzymes only react with single stereoisomers of CPP (Cho et al., 2004; Nemoto et al., 2004; Otomo et al., 2004a; Peters et al., 2000; Wilderman et al., 2004).

Prototypical plant terpene synthases are similar in size and consist of two structurally defined domains that have been simply termed the N- and C-terminal domains because these are associated with the corresponding region of their polypeptide sequence (Starks et al., 1997; Whittington et al., 2002). While both class II and class I labdane-related diterpene synthases are phylogenetically related to other plant terpene synthases (Bohlmann et al., 1998; Martin et al., 2004), these also invariably contain additional N-terminal sequence (~210-240 amino acids) that has been termed the 'insertional' element. However, this sequence element was almost certainly present in the ancestral terpene synthase (Bohlmann et al., 1998; Martin et al., 2004; Trapp and Croteau, 2001), and is particularly well conserved, along with the central region that corresponds to the prototypical N-terminal domain, in class II diterpene cyclases (Xu et al., 2004). Class I terpene synthases contain a DDXXD motif in their C-terminal domain that is involved in ligation of the divalent metal ion co-factors required for their diphosphate ionization-initiated reaction mechanism (Davis and Croteau, 2000). By contrast, class II diterpene cyclases contain a DXDD motif in their central domain that is required for their protonation-initiated cyclization reactions (Peters and Croteau, 2002a; Peters et al., 2001). Indeed, the C-terminal domain has been termed the class I domain, with the prototypical N-terminal or central domain termed the class II domain

(Christianson, 2006; Wendt and Schulz, 1998). Given the equally conserved nature of the 'insertional' element and central domain in class II diterpene cyclases, we suggest the use of class IIa and class IIb domains for these regions, respectively. Hence, the domain structure of plant labdane-related diterpene synthases consists of the class IIa, class IIb, and class I domains, although these all seem to be structurally interdependent and can not be divided into separate polypeptides that exhibit the associated activity (Peters et al., 2003).

Based on this characteristic domain structure and the class specific aspartate rich motifs, four class II and eleven class I labdane-related diterpene synthases have been found in the extensive sequence information available for rice. All four class II genes have been characterized, demonstrated to produce CPP, and were termed CPP synthases (CPS); specifically OsCPS1-4, with OsCPS1 and OsCPS2 producing *ent*-CPP (**2**) for gibberellin and phytoalexin biosynthesis, respectively, OsCPS3 being a pseduo-gene, and OsCPS4 producing *syn*-CPP (**3**) (Otomo et al., 2004b; Prisic et al., 2004; Sakamoto et al., 2004; Xu et al., 2004). The class I labdane-related diterpene synthases have been termed kaurene synthases (KS), specifically OsKS1-10 (Margis-Pinheiro et al., 2005; Otomo et al., 2004a; Sakamoto et al., 2004). However, mutational analysis has demonstrated that only OsKS1 is involved in gibberellin biosynthesis and, presumably, produces kaurene (Margis-Pinheiro et al., 2005; Sakamoto et al., 2004). The other characterized family members do not produce kaurene (**4**) and some also have been given alternative names (e.g. OsDTC1, OsDTC2, and OsDTS2). Furthermore, the same OsKS nomenclature has been assigned to different genes by Sakamoto et al. (2004) and Margis-Pinheiro et al. (2005). Thus, to avoid confusion we have suggested use of OsKSL (rice kaurene synthase-like), with the numbering scheme used by Sakamoto et al. (2004) where appropriate, for the non-kaurene producing class I labdane-

related diterpene synthases from rice (Morrone et al., 2006). Accordingly, OsKS1 retains its original designation, but the *syn*-pimaradiene synthase originally termed OsDTS2 (Wilderman et al., 2004) or OsKS4 (Otomo et al., 2004a) should be referred to as OsKSL4, the *ent*-cassadiene synthase originally referred to as OsDTC1 (Cho et al., 2004) is OsKSL7, the *syn*-stemarene synthase originally termed OsDTC2 (Nemoto et al., 2004) is OsKSL8, and the *ent*-sandaracopimaradiene synthase originally referred to as OsKS10 (Otomo et al., 2004a) is OsKSL10 (see Table 1). As indicated by this listing, a number of the OsKSL family members were uncharacterized. Here we report cloning and characterization of many of the OsKSL enzymes, along with more general analysis of this gene family, and comparison to a similar report that appeared during the preparation of this manuscript (Kanno et al., 2006).

2. Results

2.1 Identification of rice kaurene synthase-like genes

Because of our interest in using CPP stereospecific diterpene synthases as model systems for investigating substrate and product specificity, we undertook a functional genomics based approach towards identifying the enzymatic activity of the corresponding class I labdane-related diterpene synthases from rice. Putative synthases were identified in silico using the sequence of the kaurene synthase from *Arabidopsis thaliana* in BLAST searches probing the extensive sequence information available for rice. The *OsKSL* genes were largely cloned on the basis of this information and assigned numbers by mapping onto the genome and comparison to the locations given by Sakamoto et al. (2004).

OsKS1 (AY347876) was obtained after identification of its role in GA biosynthesis via transposon insertional mutagenesis, as previously reported (Margis-Pinheiro et al., 2005). This partial cDNA clone was 99.9% identical to a recently released sequence (AB126933), and the one observed difference presumably reflects allelic variation, as both clones are from the *japonica* subspecies (ssp.) of rice.

OsKSL2 was initially found as a predicted gene from genomic sequencing (CAE05199), and we were only able to obtain partial cDNA clones that do not cover the full open reading frame (ORF), despite extensive efforts in both ssp. *indica* and *japonica*. We have deposited the largest such partial cDNA from ssp. japonica (DQ823350).

OsKSL3 has previously been reported as a partial cDNA sequence (Margis-Pinheiro et al., 2005). However, sequencing of multiple cDNA clones from both ssp. *indica* and *japonica*, as well as re-sequencing of the original partial cDNA clone (AY347879), conclusively demonstrated that this gene contains a single base insertion after nucleotide 1215 (DQ823351). The resulting frame-shift should result in a severely truncated protein (441 instead of 764 amino acid residues). Notably, whereas none of our other OsKSL clones have obvious splicing errors, all three of our OsKSL3 clones from ssp. *indica* cultivar (cv.) IR24 are mis-spliced at the border between exons 10 and 11 that joins nucleotides 1702 and 1703, suggesting degradation of this downstream splice site in ssp. *indica*.

OsKSL4 is the clone (AY616862) originally reported as OsDTS2 (Wilderman et al., 2004), which is 99.3% identical to another cDNA (AB126934) that was separately reported as OsKS4 (Otomo et al., 2004a). Presumably the few observed differences are due to inter-subspecies variation, as our OsKSL4/OsDTS2 clone is from ssp. *indica* cv. IR24 and the other is from ssp. *japonica* cv. Nipponbare.

OsKSL5 was initially found in the rice full-length cDNA sequencing project, although only as a single clone containing a two-base deletion that should result in a truncated protein, although there are two single-base insertions that occur later in the sequence (AK121446). While we obtained this clone from the Rice Genome Resource Center (www.rgrc.dna.affrc.go.jp) and verified the frame-shift mutations, partial cDNA sequences reported later (AB126935 and AY347882) indicated OsKSL5 was a functional gene, i.e. at least some alleles did not contain these frame-shift mutations. Encouraged by this, we proceeded to clone a cDNA covering the full ORF that also did not have the frame-shift mutations. This OsKSL5 cDNA was 99.0% identical to those previously reported, and the few observed differences are presumably due to the inter-subspecies variation between the ssp. *japonica* cv. Nipponbare used by others and the ssp. *indica* cv. IR24 that was the source of the clone reported here (DQ823352).

OsKSL6 was originally found among the genes predicted from the rice genome (Sakamoto et al., 2004). Two overlapping partial cDNA sequences (AB126936 and AY347881) that together span the entire predicted ORF were later reported. We cloned a cDNA covering the full ORF (DQ823353). This was 99.9% identical to the previously reported partial cDNAs, and the one observed difference is presumably an allelic variant, as all sequences were from ssp. *japonica* cv. Nipponbare.

OsKSL7 was originally identified as OsDTC1 (AB089272), which was cloned from cv. BL-1 rice (Cho et al., 2004). Here we report cloning a full-length ORF for OsKSL7 from ssp. *indica* cv. IR24 (DQ823354). This clone was 99.0% identical to the BL-1 derived ORF. Again, the few observed differences presumably are a result of inter-subspecies variation.

OsKSL8 was originally identified as OsDTC2 (AB118056), which was cloned from ssp. *japonica* cv. Nipponbare rice (Nemoto et al., 2004). We cloned two overlapping partial cDNAs, also from ssp. *japonica* cv. Nipponbare, which together covered the corresponding ORF. These were spliced together to (re)create an identical copy of the full-length ORF for OsKSL8.

OsKSL9 was originally reported to be a pseduogene (Sakamoto et al., 2004). In particular, the genomic sequence data indicates that the corresponding locus (Os11g28500) only encodes a partial ORF, which was not pursued.

OsKSL10 was initially found in the rice full-length cDNA sequencing project, although only as a single clone reported as containing two neighboring single-base insertions that would result in a truncated protein (AK072461). We also obtained this clone from the Rice Genome Resource Center. However, re-sequencing of this full-length cDNA demonstrated that the frame-shift insertions were a result of sequencing errors, along with two mis-called bases in the same region. This verified sequence (DQ823355) was 99.8% identical to a partial cDNA (AB126937) reported as OsKS10 (Otomo et al., 2004a). Both clones were from ssp. *japonica* cv. Nipponbare, and the very few observed differences are presumably a result of allelic variation.

OsKSL11 (DQ100373) is a full-length cDNA that has been previously reported from both ssp. *indica* and *japonica* (Morrone et al., 2006). Notably, corresponding sequence still cannot be located in the currently available rice genome (e.g. at www.gramene.org).

2.2 Functional characterization of the rice kaurene synthase-like gene products

All of the OsKS(L) proteins were expressed as fusions to the C-terminus of gluthathione-*S*-transferase (GST-OsKSL), which provided a convenient affinity tag for single step purification. When activity was not detected in the context of GST fusion proteins, it was necessary to turn to a thioredoxin fusion construct to obtain active recombinant preparations (i.e. for OsKSL8). The resulting recombinant fusion proteins were assayed with GGPP (**1**), *ent*-CPP (**2**), or *syn*-CPP (**3**) as substrate, and their enzymatic activity assessed by GC-MS analysis of organic extracts of these reactions. Enzymatic products were identified by GC-MS based comparison to authentic samples (Fig. 2). All of the active enzymes were specific for either *ent*- (**2**) or *syn*-CPP (**3**), and did not react with the alternative stereoisomer, nor with GGPP (**1**).

As expected from its known role in gibberellin biosynthesis (Margis-Pinheiro et al., 2005; Sakamoto et al., 2004), OsKS1 specifically reacts with *ent*-CPP (**2**) to produce kaur-16-ene (**4**) (Fig. 2A-C). OsKSL3 contains a frame-shift mutation and appears to be a pseudo-gene. Consistent with this interpretation, simply correcting the frame-shift does not restore activity, suggesting that other deleterious mutations have arisen. As previously demonstrated, OsKSL4 specifically utilizes *syn*-CPP (**3**) to produce pimara-7,15-diene (**5**) (Otomo et al., 2004a; Wilderman et al., 2004). Interestingly, both OsKSL5 and OsKSL6 specifically react with *ent*-CPP (**2**) to produce (iso)kaur-15-ene (**6**) (Fig. 2D-G). As previously reported for OsKSL7 (Cho et al., 2004), our ssp. *indica* variant specifically utilizes *ent*-CPP (**2**) to produce cassa-12,15-diene (**7**) (data not shown). Unlike all the other OsKS(L), which each essentially produce only a single diterpene, it has been previously reported that OsKSL8 specifically utilizes *syn*-CPP (**3**) to produce stemar-13-ene (**8**) as its major product (~70% of

the total), along with significant (~20%) production of another, unidentified diterpene (Nemoto et al., 2004). Using our clone we observed the same product mix and were able to identify the unknown diterpene as stemod-12-ene (**10'**) (Fig. 2H-L). As previously reported for OsKSL10 (Otomo et al., 2004a), our allelic clone specifically reacts with *ent*-CPP to produce (sandaraco/iso)pimara-8(14),15-diene (**9**) (data not shown). Finally, as we have previously demonstrated, OsKSL11 produces stemod-13(17)-ene (**10**) from *syn*-CPP (**3**) (Morrone et al., 2006).

2.3 Analysis of inducible rice labdane-related diterpene synthase gene promoters

A number of OsKSL enzymes are involved in phytoalexin biosynthesis (Fig. 1), and it has been reported that transcription of that specific subset of the OsKSL family, i.e. OsKSL4, OsKSL7, OsKSL8, and OsKSL10, is induced by either fungal elicitor or UV-irradiation (Cho et al., 2004; Nemoto et al., 2004; Otomo et al., 2004a; Sakamoto et al., 2004; Wilderman et al., 2004). Also reported was similarly induced transcription of the phytoalexin biosynthesis associated *ent*- and *syn*-CPP synthases, OsCPS2 and OsCPS4, respectively (Otomo et al., 2004b; Prisic et al., 2004; Sakamoto et al., 2004; Xu et al., 2004). Given the observed differences in transcriptional regulation of the various rice diterpene synthase genes and their common evolutionary origin, we expected that comparison of their promoter regions might identify *cis*-acting regulatory elements, specifically potential binding sites for transcription factors involved in the observed induction of OsKSL4, OsKSL7, OsKSL8, and OsKSL10, as well as OsCPS2 and OsCPS4. Because *cis*-acting elements are generally located within a kilobase (kb) of the initial exon, our analyses focused on the 1.0 kb of sequence upstream of the *OsKS(L)* and *OsCPS* genes (except for *OsCPS4* where only

776 bases could be analyzed due to the presence of a gap in the genome sequence at this position). Due to the large number of potential binding sites that were identified, the results from searches for known transcription factor binding sites were difficult to interpret. To find sites conserved specifically within the inducible rice diterpene synthase promoters, we searched for sequence elements, on either strand, common to all the inducible *OsCPS* and *OsKSL* genes, but not found in either of the gibberellin biosynthesis associated and non-inducible *OsCPS1* and *OsKS1*, promoter regions. From this analysis we identified four 6-mer motifs: GTTTAT, GAAATT, TGCAAT, and ATATGG.

In addition to the known involvement of labdane-related diterpene synthases, evidence has been reported suggesting that two other genes may be involved in rice diterpenoid phytoalexin biosynthesis. Specifically, while rice contains five genes homologous to kaurene oxidase (KO), the first cytochrome P450 involved in gibberellin biosynthesis (Olszewski et al., 2002), only one of these (i.e. *OsKO2*) is actually involved in phytohormone metabolism (Sakamoto et al., 2004). Transcription of the two most divergent paralogs, termed KO-like (i.e. *OsKOL4* and *OsKOL5*), has been shown to be induced by either fungal elicitor or UV-irradiation, leading to the suggestion that OsKOL4 and OsKOL5 are involved in rice diterpenoid phytoalexin biosynthesis (Itoh et al., 2004). Notably, two of the specifically conserved motifs identified above, GTTTAT and GAAATT, are also found in the promoter regions of the similarly regulated OsKOL4 and OsKOL5 genes, while not being found in that of the GA biosynthesis specific and non-inducible OsKO2 paralog. By contrast, the TGCAAT motif was not found in the promoter regions of the inducible OsKOL genes, while the ATATGG motif was found in the promoter region of the gibberellin biosynthesis associated and non-inducible OsKO2, as well as that of OsKOL4 (although not

OsKOL5). Thus, it seems likely that the two somewhat related GTTTAT and GAAATT motifs are regulatory elements involved in the observed similar transcription induction profiles of OsCPS2, OsCPS4, OsKSL4, OsKSL7, OsKSL8, OsKSL10, OsKOL4, and OsKOL5. These motifs also may be useful for identification of other enzymatic genes involved in rice phytoalexin biosynthesis, diterpenoid or otherwise (i.e. the flavonoid sakuranetin).

2.4 Sequence analysis of the rice kaurene synthase-like genes/enzymes

As selected for by the in silico search strategy, all of the *OsKS(L)* genes share significant homology, 50-57% identity, to the kaurene synthase from *Arabidopsis thaliana* (AtKS). The encoded proteins also share significant homology with AtKS, exhibiting 39-47% identity. AtKS is most closely related to OsKS1, sharing 47% identity at the amino acid level. By contrast, the OsKSL enzymes are only 39-43% identical to AtKS at the amino acid level. This is consistent with conservation of kaurene synthase activity for AtKS and OsKS1, and the release of such selective pressure for the *OsKSL* genes, enabling the observed evolution of divergent function.

Alignment of the OsKS(L) family revealed 53-93% identity at the nucleotide level and 42-89% identity at the amino acid level (Fig. 3). Examining the resulting cDNA sequence based phylogenetic relationships demonstrates that the *OsKSL* genes are not broadly conserved by function (Fig. 4). For example, *syn*-CPP (**3**) and *ent*-CPP (**2**) specific OsKSL enzymes do not group together, as the *syn*-CPP (**3**) specific OsKSL4 is more closely related to the *ent*-CPP (**2**) specific OsKS1, OsKSL7, and OsKSL10 than the similarly *syn*-CPP (**3**) specific OsKSL8 and OsKSL11. There also is no clear phylogenetic relationship

among the inducible versus non-inducible *OsKSL* genes. Although the inducible *OsKSL4*, *OsKSL7*, and *OsKSL10* may share a common ancestor, these are only distantly related to the similarly regulated *OsKSL8*, being much more closely related to the un-inducible *OsKS1* instead.

Nevertheless, there are two closely related pairs of *OsKSL* genes that do have some functional characteristics in common. OsKSL5 and OsKSL6 are the two most closely related OsKSL family members, exhibiting 93.0% identity at the nucleotide level and 89.0% identity at the amino acid level, and both are *ent*-CPP (**2**) specific and produce isokaurene (**6**). Further, both are similarly regulated, as their transcription is not inducible by UV-irradiation or fungal elicitor (Otomo et al., 2004a; Sakamoto et al., 2004). OsKSL8 and OsKSL11 are similarly closely related, exhibiting 92.0% identity at the nucleotide level and 88.6% identity at the amino acid level, and both are *syn*-CPP (**3**) specific and produce biogenetically related stemarane (e.g. **8**) and stemodane (e.g. **10** and **10'**) type diterpenes, whose cyclization mechanisms can be envisioned as largely overlapping (Fig. 5). Indeed, as demonstrated here, OsKSL8 produces significant amounts of both skeletal types (Fig. 2H-L). However, whereas OsKSL8 is associated with phytoalexin biosynthesis and exhibits inducible gene transcription (Nemoto et al., 2004), the physiological role of OsKSL11 remains unclear as its transcription does not appear to be inducible (Morrone et al., 2006).

Consistent with their observed enzymatic activity, all the OsKS(L) enzymes contain the class I activity associated DDXXD divalent metal binding motif, and not the class II activity associated DXDD motif. Comparison of the OsKS(L) with KS from other plant species revealed that, unlike the class II labdane-related diterpene synthases (e.g. CPS) that are highly conserved across the class IIa and class IIb regions, but not the class I domain (Xu

et al., 2004), class I labdane-related diterpene synthases [e.g. KS(L)] are conserved across the entire mature protein sequence (Fig. 6), i.e. excluding the poorly conserved plastid-directing transit sequence found in all mono- and di-terpene synthases (Davis and Croteau, 2000). The observed conservation of the class I domain is consistent with previous suggestions that class I activity in labdane-related diterpene synthases is carried out in an active site that is essentially entirely located in this DDXXD containing structural domain (Peters et al., 2003; Peters and Croteau, 2002b; Peters et al., 2001), just as found in other terpene synthases (Christianson, 2006).

Finally, it seems worth mentioning that the observed difference between allelic copies of the various *OsKSL* genes is limited to ≥99.8% identity, while the difference between inter-subspecies orthologs is 99.0-99.3% identity. The corresponding 0.7-1.0% inter-subspecies variation in the *OsKSL* gene family approximates the 0.5% rate of occurrence for single nucleotide polymorphisms/differences observed between the non-repetitive regions of the genomic sequences of ssp. *indica* and *japonica* (Yu et al., 2002).

3. Discussion

The specificity of class I labdane-related diterpene synthases for particular stereoisomers of CPP provides a potential model system for investigating the underlying steric constraints in the active sites of these and, by extension, other terpene synthases. Such studies would be greatly assisted by the identification of closely related yet functionally distinct class I labdane-related diterpene synthases. Notably, terpene synthases are generally conserved by taxonomic origin rather than enzymatic activity (Bohlmann et al., 1998; Martin et al., 2004). Thus, in order to provide a model system for investigation of class I labdane-

related diterpene synthases, we have been engaged in a functional genomics based effort to elucidate the individual biochemical functions of each member of the rice kaurene synthase-like (*OsKSL*) gene family. While the biochemical function of some of the corresponding enzymes have been previously reported, a number remained uncharacterized. Here we report the completion of functional characterization of essentially all the individual family members (Table 1), which further enabled analysis of this gene family more generally.

One notable finding from our earlier studies was the functional pairing of the gene encoding the *syn*-CPP synthase *OsCPS4* near that for the *syn*-pimaradiene synthase (*OsKSL4*) in the rice genome (Wilderman et al., 2004). We also noted at that time that the genes encoding phytoalexin associated *ent*-CPP synthase *OsCPS2* and the *ent*-cassadiene synthase *OsKSL7* were similarly clustered together, along with *OsKSL5* and *OsKSL6*, and hypothesized that the enzymes encoded by these latter two gene also would prove to be specific for *ent*-CPP (**2**), which has now been verified (Fig. 2D-G). Thus, the two gene clusters containing both class I and II labdane-related diterpene synthase are grouped by biochemical function and not co-regulation. In particular, each cluster represents a functional biosynthetic module encoding enzymes that act consecutively to produce and then use a specific stereoisomer of CPP (**2, 3**), but the genes are not necessarily regulated in the same manner (i.e. while transcription of *OsCPS2* and *OsKSL7* is inducible, that of *OsKSL5* and *OsKSL6* in the same gene cluster is not).

In addition to these two class I and II labdane-related diterpene synthase containing gene clusters, the other mapped *OsKSL* genes cluster together. *OsKS1*, *OsKSL2*, and *OsKSL3* are group together in a tandem array, and *OsKSL8* is found next to the partial pseudo-gene *OsKSL9* (Sakamoto et al., 2004). Further, a similar partial gene sequence for another class I

labdane-related diterpene synthase pseudo-gene is located near *OsKSL10* (rice gene locus Os12g30800). Hence, *OsKSL4* appears to be the only *OsKSL* not located near another such gene in the rice genome. While *OsKSL11* still can not be found in the currently available rice genome, there is a 'gap' in the genome sequence near *OsCPS4* and *OsKSL4*, and it is tempting to speculate that *OsKSL11* resides in this 'gap', which would be consistent with clustering of the *OsKSL* genes, as well as functional clustering by enzymatic stereospecificity (i.e. the shared *syn*-CPP (**3**) specific nature of OsCPS4, OsKSL4, and OsKSL11). In any case, it has been noted that these gene clusters all contain retrotransposon-like elements, which has been suggested to underlie the observed extensive duplication of *CPS* and *KSL* genes in rice (Sakamoto et al., 2004), and presumably also enabled their assembly into the functional biosynthetic modules noted above.

Intriguingly, it has been recently reported that barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*) similarly contain multiple copies of *CPS*- and *KS*-like genes (Spielmeyer et al., 2004). This was based on chromosomal mapping using barley gene probes, which mapped to regions of the barley and wheat genomes that are syngenic with the rice labdane-related diterpene synthase gene clusters, indicating that these gene clusters were assembled prior to the evolutionary divergence between barley, wheat, and rice. Furthermore, a number of the barley genes mapped onto the rice genome most strongly to rice genes now known to be involved in phytoalexin biosynthesis (e.g. *OsCPS4* and *OsKSL4*), additionally suggesting that labdane-related diterpenoid secondary metabolism is widespread in the grass/cereal family (Poaceae). Consistent with this hypothesis, maize (*Zea mays*) has been shown to produce labdane-related diterpenes in response to fungal infection (Mellon and West, 1979), and phylogenetic analysis of the *CPS* gene family within the Poaceae has been used to

suggest early *CPS* gene duplication and evolution of labdane-related diterpenoid biosynthesis in cereal crop plants (Harris et al., 2005; Prisic et al., 2004).

Given the number of genes examined here it is perhaps not surprising that evidence can be found for the on-going process of evolution in the rice KS(L) gene family. For example, *OsKSL3* appears to be a pseudo-gene undergoing gene 'decay'. The corresponding cDNA clearly contains a frame-shifting single base insertion that severely truncates the encoded polypeptide. Furthermore, this sequence appears to have accumulated a number of other deleterious mutations, as simply restoring the 'correct' reading frame does not result in the production of an active class I labdane-related diterpene synthase, and the observed downstream mis-splicing of the corresponding cDNA sequences from ssp. *indica* is a further indication of gene 'decay', at least in this sub-species.

The physiological role(s) of the isokaurene (**6**) produced by OsKSL5 and OsKSL6, as well as the stemodene (**10**) produced by OsKSL11, is not clear. Nevertheless, it seems likely that these diterpene hydrocarbons will be elaborated to more complex diterpenoid natural products that may have significant biological function(s). Given the important role in plant defense that the other non-gibberellin labdane-related diterpenoids play in rice and the non-inducible nature of the associated cyclases, one plausible role for isokaurene- and stemodene-derived diterpenoids would be to act as phytoanticipins and/or allelochemicals. An interesting alternative is that one or more such labdane-related diterpenoid could act as distinct (i.e. non-gibberellin) signaling molecule(s), analogous to the recently reported role of labda-11,13-diene-8α,15-diol in tobacco (Seo et al., 2003).

As a final note, while this manuscript was in preparation a similar study was published (Kanno et al., 2006), which reported identical findings regarding *OsKSL3* (i.e. this is a pseudo-gene containing a frame-shifting single base insertion) and *OsKSL6* (i.e. this encodes an *ent*-(iso)kaur-15-ene synthase). However, Kanno et al. (2006) do not report gene promoter analysis, nor do they discuss KS(L) gene evolution. There also is one significant difference. Whereas our ssp. *indica* variant of *OsKSL5* (*OsKSL5i*) encodes an *ent*-isokaurene synthase (Fig. 2D-G), their ssp. *japonica* variant of *OsKSL5* (*OsKSL5j*; AB126935) encodes an *ent*-pimara-8(14),15-diene synthase, which represents deprotonation of the probable pimarenyl^{+} intermediate in cyclization of *ent*-CPP (**2**) to kaurane type diterpenes (Fig. 7). This difference in product outcome represents an intriguing example of sub-species specific difference in natural products metabolism, although the corresponding physiological significance is not clear as there are no currently known rice metabolites derived from either diterpene. Regardless, such emerging functional divergence in product outcome, at least in ssp. *japonica*, between the closely related and, therefore, presumably recently duplicated *OsKSL5* and *OsKSL6* is consistent with the hypothesis that there is selective pressure to differentiate similarly regulated genes, and serves as a further example of the continuing process of gene evolution in the KS(L) gene family.

OsKSL5i and OsKSL5j share 98.1% identity at the amino acid level, and of the fifteen differences between their sequences, only three appear in or near the active site in modeled structures, and only two of these are conserved between OsKSL5i and OsKSL6, which both produce isokaurene. Because it has been demonstrated that very small numbers of substitutions in the active site of class I terpene synthases are sufficient to alter product profile (Greenhagen et al., 2006; Kollner et al., 2004; Yoshikuni et al., 2006), it seems likely

that these two or three changes in active site residues between OsKSL5i and OsKSL5j (i.e. Val661, Ile664, and Ile718 in OsKSL5i, which are Leu, Thr, and Val, respectively, in OsKSL5j) may be responsible for the observed difference in product outcome (i.e. abortive production of pimara-8(14),15-diene (**11**) from *ent*-CPP (**2**) by OsKSL5j, rather than the more complex cyclization to isokaurene (**6**) catalyzed by OsKSL5i), which will be used to guide future structure-function investigations.

4. Conclusions

In summary, we report here cloning and/or biochemical characterization of seven members of the rice kaurene synthase-like gene family (Table 1). While some of these previously had been analyzed either in vitro or by in planta mutagenesis, we verified here the expected production of *ent*-kaurene (**4**) by the gibberellin biosynthesis associated OsKS1, and have clarified the product composition resulting from OsKSL8 catalyzed cyclization (Fig. 2). In addition, novel data are reported demonstrating that *OsKSL3* is a pseudo-gene, while *OsKSL5i* and *OsKSL6* encode active enzymes that produce isokaurene (**6**) from *ent*-CPP (**2**). With essentially all of the rice kaurene synthase-like genes characterized, it also was possible to carry out analysis of this gene family more broadly, revealing a pair of potential defense-related inducible promoter motifs, and enabling more detailed examination of the labdane-related diterpene synthase gene clusters found in the rice genome. This particularly included examination of related gene clusters, and associated diterpene metabolism, in other species from the grass/cereal family, suggesting the widespread occurrence, and continuing evolution, of defensive labdane-related diterpenoid secondary

metabolism throughout this important crop plant family. Notably, such widespread retention of this type of defensive metabolism indicates that labdane-related diterpenes are particularly good scaffolds for assembling antibiotic compounds and further suggests that these natural products will prove to be important components of the defense response in many, if not all, cereal crop plants.

5. Experimental

5.1 General chemicals

Unless otherwise noted, all chemicals were purchased from Fisher Scientific (Loughborough, Leicestershire, UK) and molecular biology reagents from Invitrogen (Carlsbad, CA, USA). Prior to use CH_3CN and DMF were distilled from P_2O_5 and stored over 3 Å molecular sieves.

5.2 Instrumentation

All NMR spectroscopy was carried out in the School of Chemical Sciences NMR facility at the University of Illinois. ^{1}H NMR spectra in $CDCl_3$ and CD_3OD were referenced internally with $CHCl_3$ (7.26 ppm) or with CD_3OH (4.78 ppm). ^{31}P NMR spectra was carried out in CD_3OD with 85% H_3PO_4 as external reference (0.00 ppm). Gas chromatography-mass spectrometry (GC-MS) analyses were performed using an HP1-MS column on an Agilent (Palo Alto, CA, USA) 6890N GC instrument with a 5973N mass selective detector located in the W.M. Keck Metabolomics Research Laboratory at Iowa State University, as previously described (Xu et al., 2004). Briefly, 5 μL samples were injected at 40°C in splitless mode,

the oven temperature held for 3 min., then raised at 20°C/min. to 300°C, and held there for 3 min. MS data were collected from 50 to 500 *m/z* during the temperature ramp.

5.3 Diterpene substrates and standards

GGPP (**1**) was purchased from Sigma-Aldrich. The acquisition of reference samples of sandaracopimaradiene, *ent*-kaurene (**4**), *syn*-pimaradiene (**5**), and *syn*-stemarene (**8**), along with the preparation of *syn*-CPP (**3**), have been previously described (Mohan et al., 1996). A reference sample of *ent*-isokaurene (**6**) was prepared by separation of a 1:1 mixture of *endo*- and *exo*-cyclic isomers (**6** and **4**, respectively) obtained by I_2 equilibration of *ent*-kaurene (**4**) (Bell et al., 1966) using $AgNO_3$-silica gel chromatography (Williams and Mander, 2001). The *exo*- and *endo*- isomers of *syn*-stemodene (**10** and **10'**, respectively) have also been previously described (White and Somers, 1994), and later separated and characterized (Morrone et al., 2006). A reference sample of *ent*-cassadiene (**7**) was kindly provided by Drs. Arata Yajima and Goro Yabuta (Yajima et al., 2004).

5.4 Synthesis of ent-CPP

While the preparation of *ent*-CPP (**2**) has been previously described (Mohan et al., 1996), it was prepared for this study using a modified procedure. For this process *ent*-copalol (0.73 g) was obtained from Brazilian copal resin as previously outlined (Mohan et al., 1996). The modified procedures for conversion of *ent*-copalol to the diphosphate is presented below.

Ent-copalyl chloride: The preparation of *ent*-copalyl chloride followed the method of Collington and Meyers (Meyers and Collington, 1971). A solution of *ent*-copalol (0.050 g, 0.171 mmol) and 2,4,6-collidine (0.208 g, 1.71 mmol) under N_2 was stirred and cooled at 0°C as a suspension of LiCl (0.073 g, 1.712 mmol) partially dissolved in dry DMF (3.0 mL) was added. CH_3SO_2Cl (0.059 g, 0.514 mmol) was added dropwise. After 1.5 h at 0 °C, the pale yellow reaction mixture was poured into ice-water (10 mL), and the product was extracted with cold ether (4 x 20 mL). The combined organic phases were washed with saturated aqueous $Cu(NO_3)_2$ (2 x 10 mL), 5 % $NaHCO_3$ (2 x 10 mL) and brine (2 x10 mL), dried ($MgSO_4$), and evaporated to yield a colorless oil (0.050 g; 95% yield), which was used without further purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 5.40 (tm, *J* = 8.1 Hz, 1 H, =C*H*CH_2), 4.81 (br s, 1 H, =CH_2), 4.48 (br s, 1 H, =CH_2), 4.09 (d, *J* = 7.9 Hz, 2 H, CH_2OPP), 2.37 (ddd, *J* = 12.7, 4.0, 2.4 Hz, 1 H), 2.16 (tt, *J* = 9.9, 4.3 Hz, 1 H), 1.95 (td, *J* = 13.0, 4.9 Hz, 1 H), 1.80-1.88 (m, 1 H), 1.73-1.75 (m, 1 H), 1.71 (s, 3 H, CH_3C=), 1.53-1.57 (m, 2 H), 1.44-1.52 (m, 1 H), 1.38-1.40 (m, 1 H), 1.35-1.37 (m, 1 H), 1.27-1.33 (m, 1 H), 1.24-1.26 (m, 1 H), 1.15-1.20 (m, 1 H), 1.06 (dd, *J* = 12.5, 2.6 Hz, 1 H), 0.95-1.02 (m, 1 H), 0.85 (s, 3 H), 0.78 (s, 3 H), 0.66 (s, 3 H) ppm; ^{13}C NMR (100 MHz, CD_3OD) δ 148.3, 143.5, 124.3, 106.1, 55.8, 55.3, 41.9, 41.1, 38.8, 38.0, 33.4, 30.1, 24.2, 21.7, 21.5, 19.9, 16.0, 14.3 ppm.

Ent-CPP: The conversion of *ent*-copalyl chloride to *ent*-CPP (**2**) was modeled after two methods already in the literature (Mohan et al., 1996; Woodside et al., 1993) using previously described ion exchange and purification methods (Zhao, 2005). In a dry vial equipped with a stirring bar were placed *ent*-copalyl chloride (0.050 g, 0.162 mmol) and dry 4 Å molecular sieves (0.70 g). Dry CH_3CN (2.0 mL) and HOPP $(NBu_4)_3$ (0.296 g, 0.324

mmol) were added quickly. The vial was sealed with a septum and purged with N_2, and the suspension was stirred at room temp for 12 h. The solids were removed by filtration and washed with CH_3CN (70 mL). The CH_3CN filtrate was washed with pentane (3 x10 mL), and CH_3CN was removed with a rotary evaporator. The remaining tetrabutylammonium salt was dissolved in 18 mL of 0.2 % (w/v) NH_4HCO_3 and 2 % (v/v) isopropyl alcohol in H_2O (solvent A) and applied to a Bio-Rad AG50W-X8 ion exchange column (d 1.2 cm, h 25 cm, volume 30 mL) that had previously been freshly washed with saturated NH_4OH and deionized water until pH = 7.0, and the column was eluted with solvent A (42 mL). The total eluate (~60 mL) was lyophilized 3 times for 3 h each. The crude white powder was suspended in dry MeOH (8 mL), vortexed, and centrifuged, then the supernatant removed and concentrated to 0.5-0.6 mL in vacuo. This extraction/vortexing/centrifugation/concentration procedure was repeated 3 times. ^{31}P NMR analyses of the 4 supernatant fractions were conducted to determine the ratio of product diphosphate (2d, δ_P -8 to -9 ppm) to PPi (s, δ_P +3 ppm). Supernatant fractions 1 and 2 were combined and concentrated to give *ent*-CPP (**2**) NH_4 salt as a white powder, which was not further purified (yield: 17.6 mg, purity 82%), in 19% overall yield for the 2 steps from *ent*-copalol. ^{31}P NMR (376 MHz, CD_3OD) δ -8.25 (d, J = 16.6 Hz, 1 P), -8.85 (d, J = 19.3 Hz, 1 P) ppm; ^{1}H NMR (400 MHz, CD_3OD) δ 5.37 (tm, J = 8.0 Hz, 1 H, =C*H*CH_2), 4.83 (s, 1 H, =C*H*$_2$), 4.52 (s, 1 H, =C*H*$_2$), 3.35 (br s, 2 H, CH_2OPP), 2.39 (ddd, J = 12.5, 4.0, 2.4 Hz, 1 H), 2.11-2.18 (m, 1 H), 1.95-2.01 (m, 1 H), 1.82-1.87 (m, 1 H), 1.73-1.80(m, 1 H), 1.69 (s, 3 H, CH_3C=), 1.57-1.64 (m, 2 H), 1.46-1.51 (m, 1 H), 1.40-1.44 (m, 1 H), 1.36-1.40 (m, 1 H),

1.33 (dd, *J* = 13.0, 4.3 Hz, 1 H), 1.26-1.30 (m, 1 H), 1.17-1.25 (m, 1 H), 1.12 (dd, *J* = 12.7, 2.6 Hz, 1 H), 1.00-1.07 (m, 1 H) ppm.

5.5 Plant material

Rice plants (*O. sativa* L. ssp. *indica* cv. IR24), and seedlings (ssp. *japonica* cv. Nipponbare) were those previously described (Xu et al., 2004). Briefly, leaves were detached from four-week old greenhouse grown plants and UV-irradiated for 25 min. from 15 cm. with 254 nm wavelength light. These were then incubated for 24 hrs in dark humid conditions at 30°C. Seedlings were germinated from surface sterilized seeds on filter paper on top of 1.2% agar plates (8 seeds per plate) incubated in the dark at 30°C for one week. The seedlings were then sprayed with approximately 2 mL 0.1% Tween 20 containing 0.5 mM methyl jasmonate per plate and incubated under the same conditions for 24 hrs. Total RNA was extracted using the Concert Plant RNA Reagent, and mRNA purified using Dynabeads Oligo$(dT)_{25}$ (Dynal Biotech, Oslo, Norway).

5.6 Cloning

The OsKSL genes were largely obtained via RT-PCR using the primer pairs shown in Table 2. All of the genes were initially cloned into pCR-BluntII-TOPO and completely sequenced. The corresponding ORFs were then transferred to the Gateway vector system via PCR amplification and directional topoisomerization based insertion into pENTR/SD/D-TOPO and verified by complete sequencing. The resulting clones were subsequently transferred via directional recombination to the T7-based N-terminal GST fusion expression

vector pDEST15 and, for OsKSL8, the T7-based N-terminal thioredoxin fusion expression vector pTH8 (Hammarström et al., 2002).

For OsKS1 we used the partial cDNA (AY347876) previously reported as OsKS1A (Margis-Pinheiro et al., 2005). For OsKSL2, despite repeated attempts with both RT-PCR and RACE reactions from both ssp. *japonica* cv. Nipponbare and ssp. *indica* cv. IR24, it was only possible to generate fragmentary sequence information. The largest such partial cDNA (1545 nucleotides), from ssp. *japonica* cv. Nipponbare, was deposited (DQ823350). For OsKSL3 three independent full-length cDNAs were obtained from both ssp. *japonica* cv. Nipponbare and ssp. *indica* cv. IR24 rice. All six clones contained an 'extra' adenosine base after nucleotide 1215 of the originally predicted sequence. This was further confirmed by re-sequencing of the originally reported partial cDNA clone (AY347879), which also proved to contain this frame-shift insertion. The three clones from ssp. *indica* cv. IR24 were additionally mis-spliced between exons 10 and 11 (joining nucleotides 1702 and 1703), with one being un-spliced (i.e. containing the corresponding intron) and the other two missing nucleotides 1703-1709 (i.e. as a result of being spliced at the wrong nucleotide). Outside of these splicing differences the six full-length clones were approximately 99% identical, with almost all of the differences arising from inter-subspecies variation between the ssp. *indica* and ssp. *japonica* clones. A representative cDNA clone from ssp. *japonica* cv. Nipponbare has been deposited (DQ823351). The frame-shift insertion in this particular cDNA was also 'removed' by PCR-based mutagenesis to generate a 'corrected' OsKSL3. Cloning of OsKSL4 from ssp. *indica* cv. IR24 has been previously reported (Wilderman et al., 2004). For OsKSL5 two independent clones were obtained from ssp. *indica* cv. IR24 and fully sequenced, demonstrating complete identity, to verify the deposited sequence (DQ823352).

For OsKSL6 two independent clones were obtained from ssp. *japonica* cv. Nipponbare and fully sequenced, demonstrating complete identity, to verify the deposited sequence (DQ823353). For OsKSL7 two independent clones were obtained from ssp. *indica* cv. IR24 and fully sequenced, demonstrating complete identity, to verify the deposited sequence (DQ823354). For OsKSL8 we were not able to clone a full-length cDNA, instead consistently finding the closely related OsKSL11 in both ssp. *indica* and *japonica* (Morrone et al., 2006). It was possible to clone partial cDNA fragments separately corresponding to the first 1715 and last ~900 nucleotides of the reported ORF of OsKSL8 (AB118056; 2463 nucleotides total). However, sequencing of the 3' end fragment revealed a 21 nucleotide deletion in the last exon (i.e. nucleotides 2171-2192 were missing), which is extremely GC-rich and contains quite repetitive sequence elements in this region. Attempts to clone a 3' end fragment with the correct sequence were unsuccessful, and ultimately the missing nucleotides were inserted via PCR extension, and the full-length ORF subsequently re-created using the 5' end and corrected 3' end fragments in a mega-primed PCR approach. For OsKSL10 a full-length cDNA was obtained from the Rice Genome Resource Center (ssp. *japonica* cv. Nipponbare), re-sequenced, and the corrected sequence deposited (DQ823355). Cloning of OsKSL11 from both ssp. *indica* and *japonica* has been previously reported (Morrone et al., 2006).

5.7 Recombinant expression

Heterologous expression was performed using the OverExpress C41 strain of *E. coli* (Avidis, France), as previously described (Xu et al., 2004). Briefly, 50-mL NZY cultures inoculated from 5-7 individual transformants/colonies were grown at 37°C to mid-log phase

(A_{600} ~ 0.6), then transferred to 16°C for 1-2 hrs prior to induction with 1 mM IPTG for overnight (14-18 hrs) expression. Cells were harvested by centrifugation, resuspended in 1 mL cold lysis buffer (50 mM Bis-Tris, pH 6.8, 1 mM DTT), lysed by mild sonication (15 sec., continuous output, half-maximum power), and clarified by centrifugation (40,0000*g*, 20 min.). Recombinant GST-tagged proteins were purified using GST-agarose beads (Sigma-Alrich, St. Louis, MO, USA) in batch mode. The thioredoxin-OsKSL8 fusion protein was only partially purified using ceramic hydroxyapetite type II (BioRad, Hercules, CA, USA) in batch mode. The batch purification was carried out at 4°C during which clarified extract was incubated 5-10 min. with ~0.3 mL beads, washed 5 times with 1 mL lysis buffer, then eluted in 0.5 mL (either 10 mM glutathione in lysis buffer for GST-OsKSL bound to glutathione beads, or 0.2 M sodium phosphate, pH 6.8, for thioredoxin-OsKSL8 bound to ceramic hydroxyapetite), the resulting eluate was filtered (0.2 μ) and immediately used in enzymatic assays.

5.8 Functional characterization

Assays were generally carried out with GGPP (5 μg) as substrate in coupled reactions with either the *syn*-CPP synthase OsCPS4 or *ent*-CPP synthase ZmCPS2/An2 (Harris et al., 2005), both expressed and purified as GST-fusion proteins. For every gene product individual enzyme assays also were run with 5 μg GGPP (**1**), *ent*-CPP (**2**), or *syn*-CPP (**3**) directly. Reactions were run in 0.5 mL assay buffer (50 mM Hepes, pH 7.2, 10 mM $MgCl_2$, 10% glycerol, and 5 mM DTT) using 25 μL recombinant CPS and/or 25 μL recombinant OsKS(L) for 3-16 hrs at room temperature. These assays were extracted 3 times with an equal volume of hexanes, which were pooled, dried under nitrogen, and re-dissolved in 100 μL of hexanes prior to GC-MS analysis.

5.9 Sequence analysis

BLAST searches were carried out on-line at both GenBank (www.ncbi.nih.gov) and TIGR (www.tigr.org). The cloned cDNAs were mapped back onto the rice genome by web-based alignment at Gramene (www.gramene.org), and the corresponding up-stream sequences also obtained here. Notably, 776 bases upstream of the OsCPS4 gene there seems to be a gap in the genome sequence, as represented by a stretch of several hundred unidentified bases (i.e. 'N's). Promoter analysis was done on-line at PlantCare (bioinformatics.psb.ugent.be/webtools/ plantcare/html) and TAIR (www.arabidopsis.org/tools/bulk/motiffinder/index.jsp). All other sequence analysis was carried out with the VectorNTI software package (Invitrogen) using standard parameters.

Acknowledgements

We thank Drs. Arata Yajima and Goro Yabuta (Tokyo University of Agriculture) for graciously providing an authentic sample of synthetic *ent*-cassa-12,15-diene, and Dr. Gustavo MacIntosh (Iowa State University) for assistance with promoter analysis. This study was generously supported by grants from the USDA-CSREES (2005-35318-15477) to R.J.P., and from the NIH (GM13956) to R.M.C., along with fellowship support from Iowa State University (to D.M.), while the work of M.M.-P. and N.M.U. in the CSIRO Plant Industry's "Rice Functional Genomics" group (http://www.pi.csiro.au/fgrttpub/home.htm) was supported by GrainGene (Australia), Rural Industries Research and Development Corporation (RIRDC), Australia, and the NSW Agricultural Genomics Centre.

References

Akatsuka, T., Kodama, O., Sekido, H., Kono, Y., Takeuchi, S., 1985. Novel phytoalexins (oryzalexins A, B, and C) isolated from rice blast leaves infected with *Pyricularia oryzae*. Part I: Isolation, characterization and biological activities of oryzalexins. Agric. Biol. Chem. 49, 1689-1694.

Bell, R. A., Ireland, R. E., Partyka, R. A., 1966. Experiments directed toward the total synthesis of terpenes. VIII. The total synthesis of (+ or -) kaurene and (±)-atisirene. J. Org. Chem. 31, 2530-2536.

Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc. Natl. Acad. Sci. USA 95, 4126-4133.

Cartwright, D. W., Langcake, P., Pryce, R. J., Leworthy, D. P., Ride, J. P., 1977. Chemical activation of host defence mechanisms as a basis for crop protection. Nature 267, 511-513.

Cartwright, D. W., Langcake, P., Pryce, R. J., Leworthy, D. P., Ride, J. P., 1981. Isolation and characterization of two phytoalexins from rice as momilactones A and B. Phytochemistry 20, 535-537.

Cho, E.-M., Okada, A., Kenmoku, H., Otomo, K., Toyomasu, T., Mitsuhashi, W., Sassa, T., Yajima, A., Yabuta, G., Mori, K., Oikawa, H., Toshima, H., Shibuya, N., Nojiri, H., Omori, T., Nishiyama, M., Yamane, H., 2004. Molecular cloning and characterization of a cDNA encoding *ent*-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor. Plant J. 37, 1-8.

Christianson, D. W., 2006. Structural biology and chemistry of the terpenoid cyclases. Chem. Rev. 106, 3412-3442.

Davis, E. M., Croteau, R., 2000. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. Top. Curr. Chem. 209, 53-95.

Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W. L., Chen, L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296, 92-100.

Greenhagen, B. T., O'Maille, P. E., Noel, J. P., Chappell, J., 2006. Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. Proc Natl Acad Sci U S A. 103, 9826-9831.

Hammarström, M., Hellgren, N., van Den Berg, S., Berglund, H., Härd, T., 2002. Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci. 11, 313-321.

Harris, L. J., Saparno, A., Johnston, A., Prisic, S., Xu, M., Allard, S., Kathiresan, A., Ouellet, T., Peters, R. J., 2005. The maize *An2* gene is induced by *Fusarium* attack and encodes an *ent*-copalyl diphosphate synthase. Plant Mol. Biol. 59, 881-894.

Itoh, H., Tatsumi, T., Sakamoto, T., Otomo, K., Toyomasu, T., Kitano, H., Ashikari, M., Ichihara, S., Matsuoka, M., 2004. A rice semi-dwarf gene, *Tan-Ginbozu* (*D35*), encodes the gibberellin biosynthesis enzyme, *ent*-kaurene oxidase. Plant Mol. Biol. 54, 533-547.

Kanno, Y., Otomo, K., Kenmoku, H., Mitsuhashi, W., Yamane, H., Oikawa, H., Toshima, H., Matsuoka, M., Sassa, T., Toyomasu, T., 2006. Characterization of a rice gene family encoding type-A diterpene cyclases. Biosci. Biotechnol. Biochem. 70, 1702-1710.

Kato, H., Kodama, O., Akatsuka, T., 1993. Oryzalexin E, a diterpene phytoalexin from UV-irradiated rice leaves. Phytochemistry 33, 79-81.

Kato, H., Kodama, O., Akatsuka, T., 1994. Oryzalexin F, a diterpene phytoalexin from UV-irradiated rice leaves. Phytochemistry 36, 299-301.

Kato-Noguchi, H., Ino, T., 2003. Rice seedlings release momilactone B into the environment. Phytochemistry 63, 551-554.

Kato-Noguchi, H., Ino, T., Sata, N., Yamamura, S., 2002. Isolation and identification of a potent allelopathic substance in rice root exudates. Physiologia Plantarum 115, 401-405.

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C. J., Ohtsuki, K., Shishiki, T., Otomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M., Hayashizaki, Y., 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice. Science 301, 376-379.

Kodama, O., Li, W. X., Tamogami, S., Akatsuka, T., 1992. Oryzalexin-S, a novel stemarane-type diterpene rice phytoalexin. Biosci. Biotechnol. Biochem. 56, 1002-1003.

Koga, J., Ogawa, N., Yamauchi, T., Kikuchi, N., Ogasawara, N., Shimura, M., 1997. Functional moiety for the antifungal activity of phytocassane E, a diterpene phytoalexin from rice. Phytochemistry 44, 249-253.

Koga, J., Shimura, M., Oshima, K., Ogawa, N., Yamauchi, T., Ogasawara, N., 1995. Phytocassanes A, B, C, and D, novel diterpene phytoalexins from rice, *Oryza sativa* L. Tetrahedron 51, 7907-7918.

Kollner, T. G., Schnee, C., Gershenzon, J., Degenhardt, J., 2004. The variability of sesquiterpenes emitted from two *Zea mays* cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. Plant Cell 16, 1115-1131.

MacMillan, J., Beale, M. H., 1999. Diterpene biosynthesis. In: Cane, D. E. (Ed.), Isoprenoids Including Carotenoids and Steroids, vol. 2. Elsevier Science Ltd., Oxford, pp. 217-243.

Margis-Pinheiro, M., Zhou, X.-R., Zhu, Q.-H., Dennis, E. S., Upadhyaya, N. M., 2005. Isolation and characterization of a *Ds*-tagged rice (*Oryza sativa* L.) GA responsive dwarf mutant defective in an early step of the gibberellin biosynthesis pathway. Plant Cell Reports 23, 819-833.

Martin, D. M., Faldt, J., Bohlmann, J., 2004. Functional characterization of nine norway spruce *TPS* genes and evolution of gymnosperm terpene synthases of the *TPS-d* subfamily. Plant Physiol. 135, 1908-1927.

Mellon, J. E., West, C. A., 1979. Diterpene biosynthesis in maize seedlings in response to fungal infection. Plant Physiol. 64, 406-410.

Meyers, A. I., Collington, E. W., 1971. Facile and specific conversion of allylic alcohols to allylic chlorides without rearrangement. J. Org. Chem. 36, 3044-3045.

Mohan, R. S., Yee, N. K. N., Coates, R. M., Ren, Y. Y., Stamenkovic, P., Mendez, I., West, C. A., 1996. Biosynthesis of cyclic diterpene hydrocarbons in rice cell suspensions: conversion of 9,10-*syn*-labda-8(17),13-dienyl diphosphate to 9ß-pimara-7,15-diene and stemar-13-ene. Arch. Biochem. Biophys. 330, 33-47.

Morrone, D., Jin, Y., Xu, M., Choi, S.-Y., Coates, R. M., Peters, R. J., 2006. An unexpected diterpene cyclase from rice: Functional identification of a stemodene synthase. Arch. Biochem. Biophys. 448, 133-140.

Nemoto, T., Cho, E.-M., Okada, A., Okada, K., Otomo, K., Kanno, Y., Toyomasu, T., Mitsuhashi, W., Sassa, T., Minami, E., Shibuya, N., Nishiyama, M., Nojiri, H., Yamane, H., 2004. Stemar-13-ene synthase, a diterpene cyclase involved in the biosynthesis of the phytoalexin oryzalexin S in rice. FEBS Lett 571, 182-186.

Olszewski, N., Sun, T.-P., Gubler, F., 2002. Gibberellin signalling: Biosynthesis, catabolism, and response pathways. Plant Cell Supplement 2002, S61-S80.

Otomo, K., Kanno, Y., Motegi, A., Kenmoku, H., Yamane, H., Mitsuhashi, W., Oikawa, H., Toshima, H., Itoh, H., Matsuoka, M., Sassa, T., Toyomasu, T., 2004a. Diterpene cyclases responsible for the biosynthesis of phytoalexins, momilactones A, B, and oryzalexins A-F in rice. Biosci. Biotechnol. Biochem. 68, 2001-2006.

Otomo, K., Kenmoku, H., Oikawa, H., Konig, W. A., Toshima, H., Mitsuhashi, W., Yamane, H., Sassa, T., Toyomasu, T., 2004b. Biological functions of *ent*- and *syn*-copalyl diphosphate synthases in rice: key enzymes for the branch point of gibberellin and phytoalexin biosynthesis. Plant J. 39, 886-893.

Peters, R. J., 2006. Uncovering the complex metabolic network underlying diterpenoid phytoalexin biosynthesis in rice and other cereal crop plants. Phytochemistry 67, 2307-2317.

Peters, R. J., Carter, O. A., Zhang, Y., Matthews, B. W., Croteau, R. B., 2003. Bifunctional abietadiene synthase: Mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. Biochemistry 42, 2700-2707.

Peters, R. J., Croteau, R. B., 2002a. Abietadiene synthase catalysis: Conserved residues involved in protonation-initated cyclization of geranylgeranyl diphosphate to (+)-copalyl diphosphate. Biochemistry 41, 1836-1842.

Peters, R. J., Croteau, R. B., 2002b. Abietadiene synthase catalysis: Mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. Proc. Natl. Acad. Sci. U.S.A. 99, 580-584.

Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. M., Lee, H.-J., Coates, R. M., Croteau, R. B., 2000. Abietadiene synthase from grand fir (*Abies grandis*): Characterization and mechanism of action of the "pseudomature" recombinant enzyme. Biochemistry 39, 15592-15602.

Peters, R. J., Ravn, M. M., Coates, R. M., Croteau, R. B., 2001. Bifunctional abietadiene synthase: Free diffusive transfer of the (+)-copalyl diphosphate intermediate between two distinct active sites. J. Am. Chem. Soc. 123, 8974-8978.

Prisic, S., Xu, M., Wilderman, P. R., Peters, R. J., 2004. Rice contains disparate *ent*-copalyl diphosphate synthases with distinct metabolic functions. Plant Physiol. 136, 4228-4236.

Sakamoto, T., Miura, K., Itoh, H., Tatsumi, T., Ueguchi-Tanaka, M., Ishiyama, K., Kobayashi, M., Agrawal, G. K., Takeda, S., Abe, K., Miyao, A., Hirochika, H., Kitano, H., Ashikari, M., Matsuoka, M., 2004. An overview of gibberellin metabolism enzyme genes and their related mutants in rice. Plant Physiol. 134, 1642-1653.

Sekido, H., Endo, T., Suga, R., Kodama, O., Akatsuka, T., Kono, Y., Takeuchi, S., 1986. Oryzalexin D (3,7-dihydroxy-(+)-sandaracopimaradiene), a new phytoalexin isolated from blast-infected rice leaves. J. Pesticide Sci. 11, 369-372.

Seo, S., Seto, H., Koshino, H., Yoshida, S., Ohashi, Y., 2003. A diterpene as an endogenous signal for the activation of defense responses to infection with *tobacco mosaic virus* and wounding in tobacco. Plant Cell 15, 863-873.

Spielmeyer, W., Ellis, M., Robertson, M., Ali, S., Lenton, J. R., Chandler, P. M., 2004. Isolation of gibberellin metabolic pathway genes from barley and comparative mapping in barley, wheat and rice. Theor. Appl. Genet. 109, 847-855.

Starks, C. M., Back, K., Chappell, J., Noel, J. P., 1997. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science 277, 1815-1820.

Trapp, S. C., Croteau, R. B., 2001. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics 158, 811-832.

VanEtten, H. D., Mansfield, J. W., Bailey, J. A., Farmer, E. E., 1994. Two classes of plant antibiotics: phytoalexins versus 'phytoanticipins'. Plant Cell 6, 1191-1192.

Wendt, K. U., Schulz, G. E., 1998. Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. Structure 6, 127-133.

White, J. D., Somers, T. C., 1994. Construction of the stemodane nucleus by a hydroxyl-directed intramolecular ene reaction. Total synthesis of (±)-2-desoxystemodinone. J. Am. Chem. Soc. 116, 9912-9920.

Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B., Christianson, D. W., 2002. Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci U S A. 99, 15375-15380.

Wickham, K. A., West, C. A., 1992. Biosynthesis of rice phytoalexins: Identification of putative diterpene hydrocarbon precursors. Arch. Biochem. Biophys. 293, 320-332.

Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., Peters, R. J., 2004. Identification of *syn*-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. Plant Physiol. 135, 2098-2105.

Williams, C. M., Mander, L. N., 2001. Chromatography with silver nitrate. Tetrahedron 57, 425-447.

Woodside, A. B., Huang, Z., Poulter, C. D., 1993. Trisammonium geranyl diphosphate. Org. Synth. Coll. Col. VIII.

Xu, M., Hillwig, M. L., Prisic, S., Coates, R. M., Peters, R. J., 2004. Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. Plant J. 39, 309-318.

Yajima, A., Mori, K., Yabuta, G., 2004. Total synthesis of *ent*-cassa-12,15-diene, a putative precursor of rice phytoalexins, phytocassanes A-E. Tetrahedron Lett. 45, 167-169.

Yoshikuni, Y., Ferrin, T. E., Keasling, J. D., 2006. Designed divergent evolution of enzyme function. Nature 440, 1078-1082.

Yu, J., Hu, S., Wang, J., Wong, G. K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, W., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., Yang, H., 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296, 79-92.

Zhao, Y., 2005. Modified farnesyl diphosphate, germacrane, and eremophilane sesquiterpenes as probes for capsidiol biosynthesis. Chemistry. University of Illinois, Urbana-Champaign.

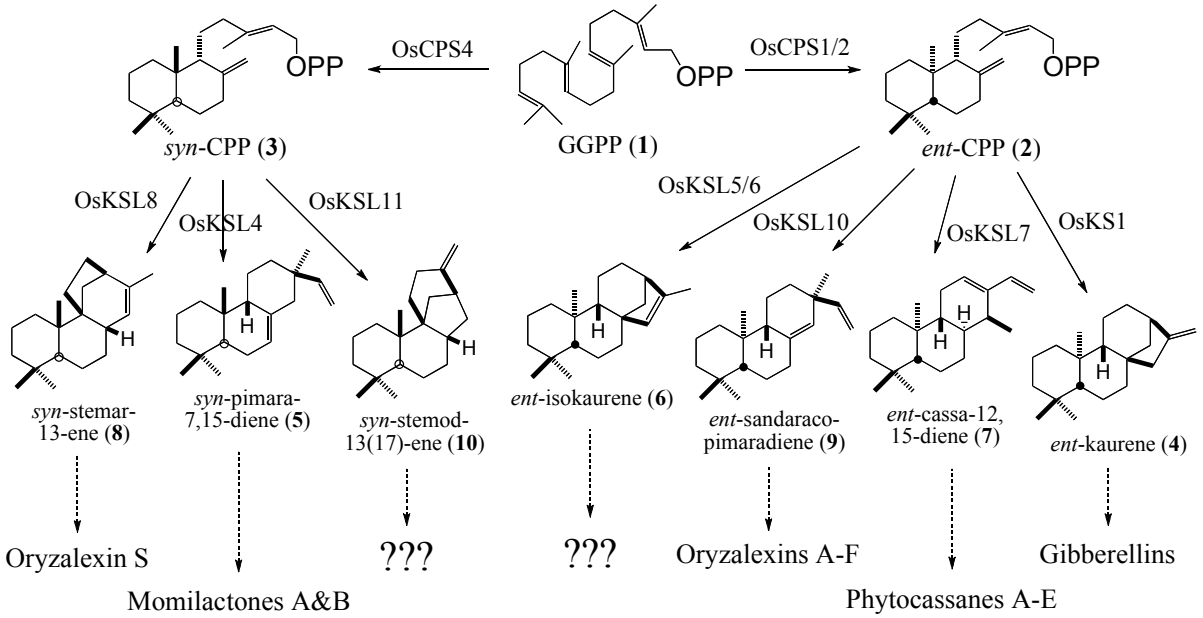


Figure 1: Known labdane-related diterpene cyclization reactions in rice. The corresponding cyclases are indicated, along with their products and, where known, the derived natural products (dashed arrows indicate multiple biosynthetic steps).

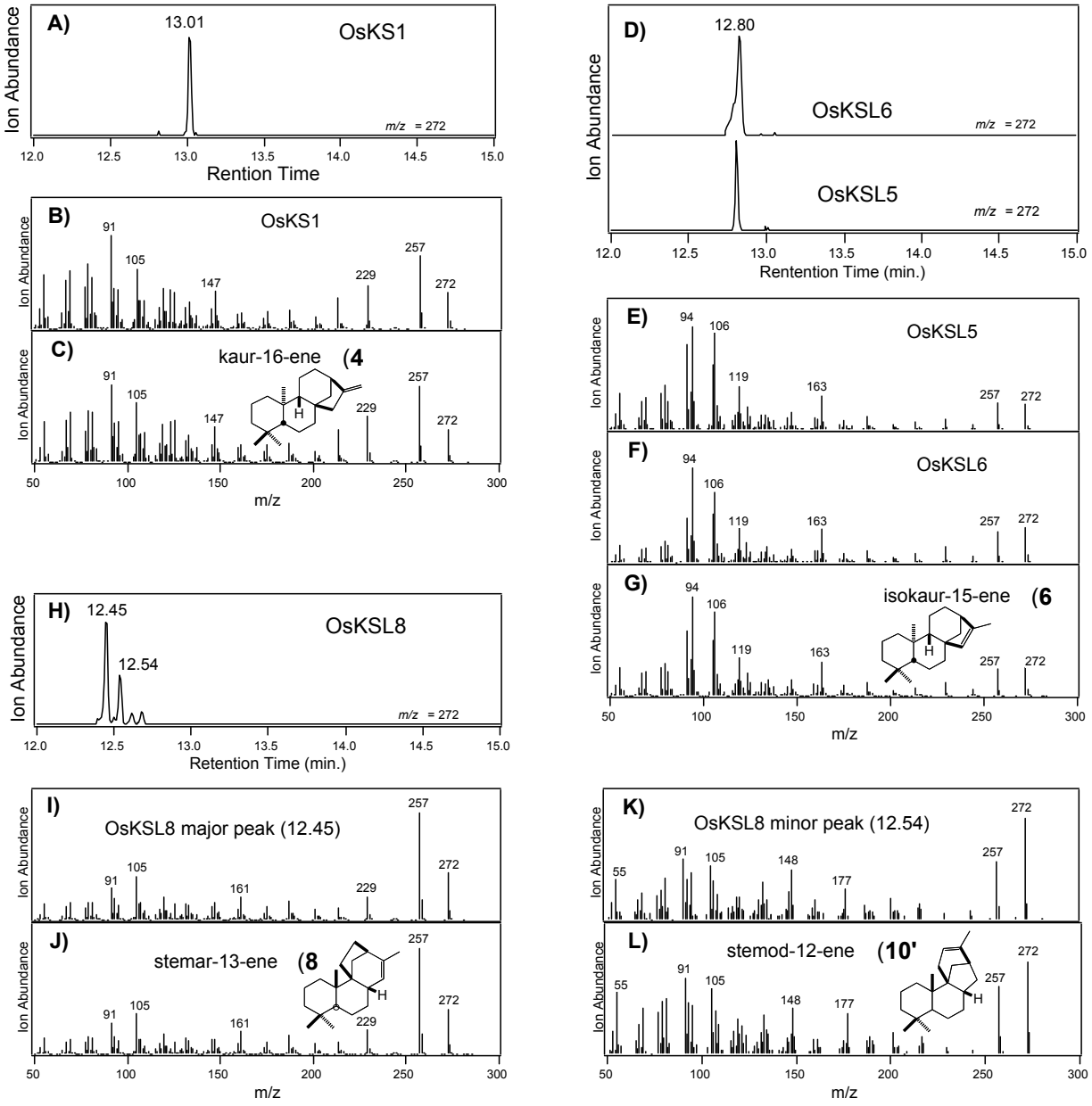


Figure 2: GC-MS analysis of products formed by selected OsKS(L). A) Chromatograph of the product formed by OsKS1 from *ent*-CPP. B) Mass spectrum of OsKS1 product peak (RT = 13.01 min.). C) Mass spectrum of authentic *ent*-kaur-16-ene (**4**) (RT = 13.01 min.). D) Chromatograph of the product formed by OsKSL5 and OsKSL6 (as indicated) from *ent*-CPP.

E) Mass spectrum of OsKSL5 product peak (RT = 12.80 min.). F) Mass spectrum of OsKSL6 product peak (RT = 12.80). G) Mass spectrum of authentic *ent*-(iso)kaur-15-ene (**6**) (RT = 12.81 min.). H) Chromatograph of the products formed by OsKSL8 from *syn*-CPP. I) Mass spectrum of the major OsKSL8 product peak (RT = 12.45 min.). J) Mass spectrum of authentic *syn*-stemar-13-ene (**8**) (RT = 12.45). K) Mass spectrum of the minor OsKSL product peak (RT = 12.54). L) Mass spectrum of authentic *syn*-stemod-12-ene (**10'**) (RT = 12.55).

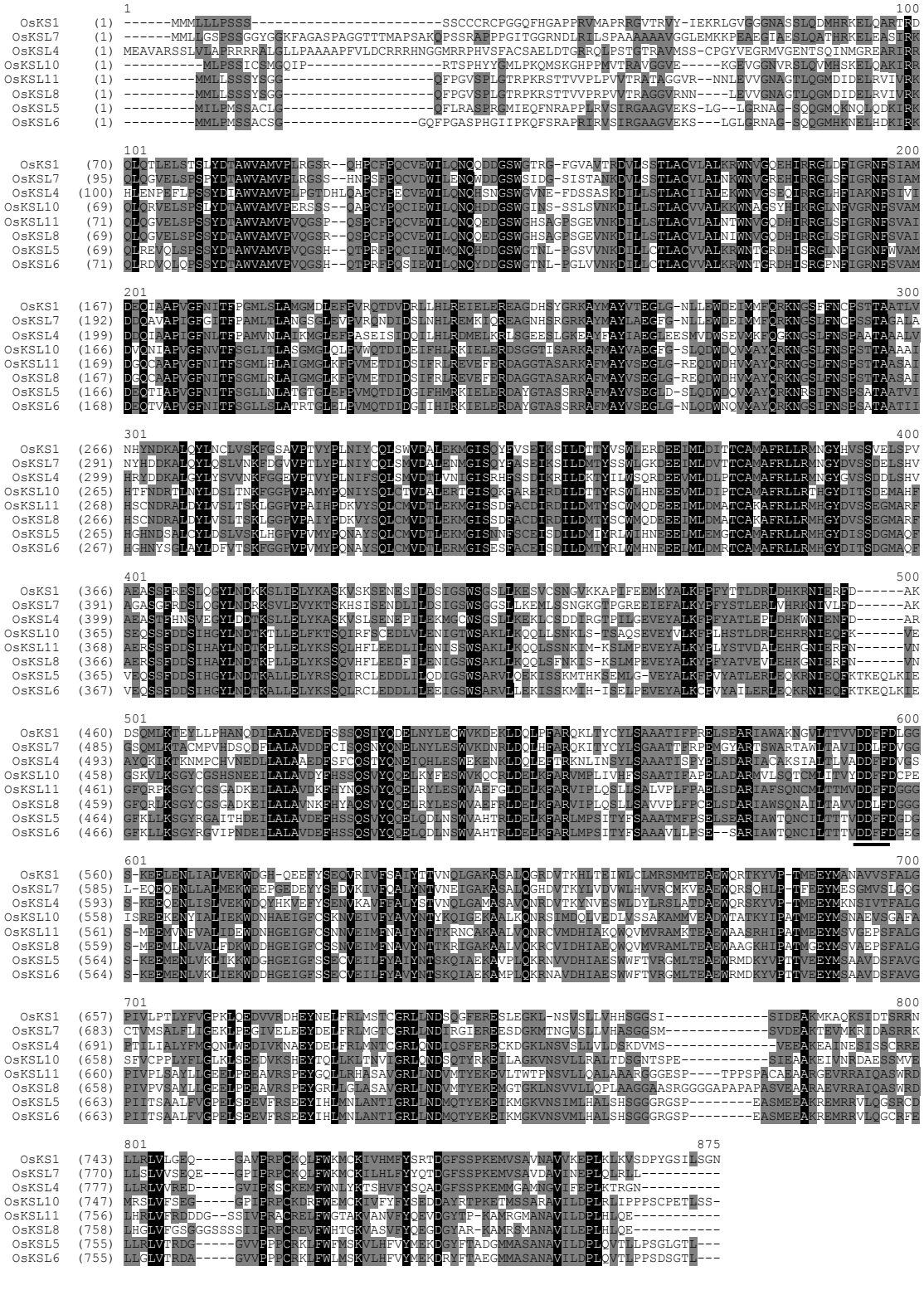


Figure 3: Sequence comparison of the rice kaurene synthase-like enzymatic family. Shown are all the active enzymes, with the DDXXD motif underlined.

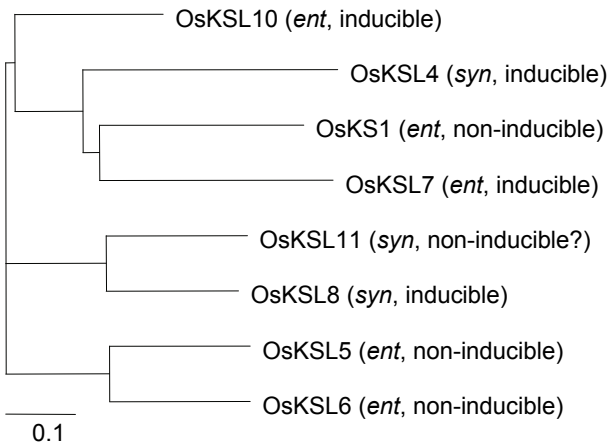


Figure 4: Phylogenetic tree illustrating the relationship of the OsKSL gene family based on alignment of the corresponding cDNA (ORF only) sequences. Indicated is the enzymatic stereospecificity and inducible or non-inducible nature of the corresponding gene transcription.

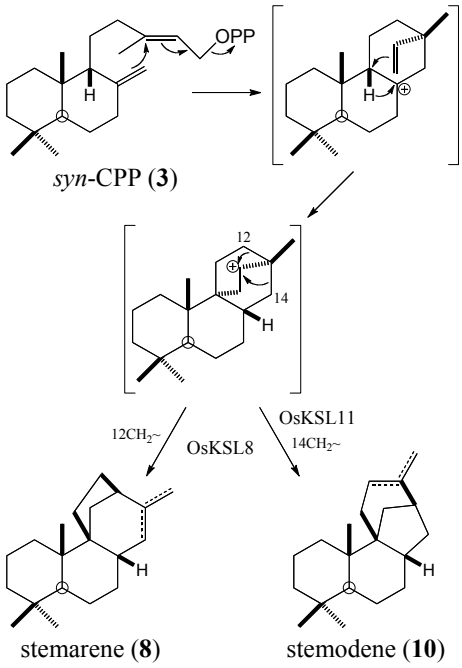


Figure 5: Proposed biogenetic cyclization mechanisms leading to formation of stemarene (**8**) and stemodene (**10**) with the product outcomes catalyzed by OsKSL8 and OsKSL11 indicated (dotted bonds indicate alternative locations for the carbon-carbon double bond).

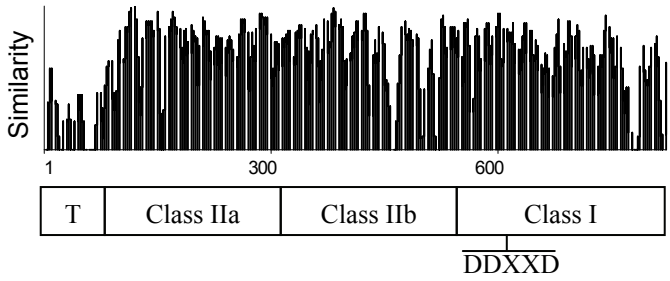


Figure 6: Histograph depicting amino acid sequence similarity over all KS(L), i.e. class I labdane-related diterpene synthases (numbering based on the consensus sequence). Also depicted is the approximate corresponding domain structure defined in the text, along with approximate location of the class I associated DDXXD motif (T indicates transit peptide region).

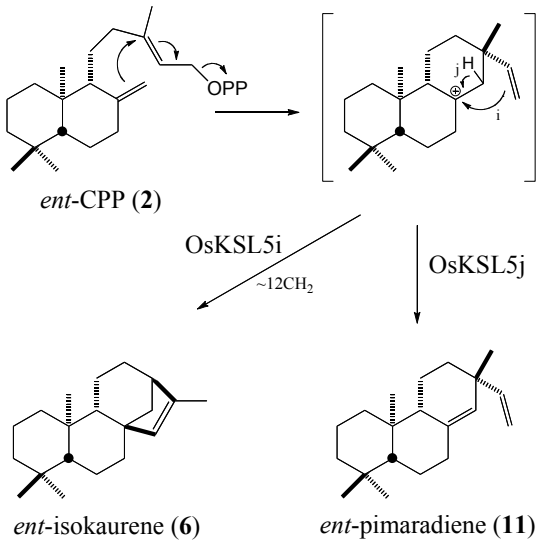


Figure 7: Proposed biogenetic cyclization mechanism catalyzed by OsKSL5i to produce iso-kaurene (**6**), along with the abortive production of pimara-8(14),15-diene (**11**) catalyzed by OsKSL5j. The further cyclization catalyzed by OsKSL5i is indicated by the arrow marked 'i', the deprotonation catalyzed by OsKSL5j is indicated by the arrow marked 'j'.

Table 1: Rice Kaurene Synthase-Like Gene Family.

Name[a]	Product	Inducible	Accession	Locus	Alternate	Reference
OsKS1	*ent*-kaurene	No	NR[b] AY347876	Os04g52230	-	Sakamoto et al. (2004) Margis-Pinheiro et al. (2005)
OsKSL2	<partial?>	???	DQ823350	Os04g52240	-	This study
OsKSL3	<pseudo-gene>	No	DQ823351	Os04g52210	OsKS2[c]	This study
OsKSL4	*syn*-pimaradiene	Yes	AY616862 AB126934	Os04g10060	OsDTS2 -	Wilderman et al. (2004) Otomo et al. (2004)
OsKSL5	*ent*-iso-kaurene	No	DQ823352	Os02g36220	OsKS6[c]	This study
OsKSL6	*ent*-iso-kaurene	No	DQ823353	Os02g36264	OsKS5[c]	This study
OsKSL7	*ent*-cassadiene	Yes	DQ823354[d]	Os02g36140	OsDTC1	Cho et al. (2004)
OsKSL8	*syn*-stemarene[e]	Yes	AB118056	Os11g28530	OsDTC2	Nemoto et al. (2004)
OsKSL9	<pseudo-gene>	-	NR[b]	Os11g28500	-	Sakamoto et al. (2004)
OsKSL10	*ent*-pimaradiene	Yes	DQ823355[d]	Os12g30824	-	Otomo et al. (2004)
OsKSL11	*syn*-stemodene	No?	DQ100373	???	-	Morrone et al. (2006)

[a]Gene numbering based on Sakamoto et al. (2004), with Kaurene Synthase-Like (KSL) nomenclature for the non-kaurene producing family members, as previously suggested (Morrone et al., 2006).

[b]NR = not reported, Sakamoto et al. (2004) did not report any sequence information.

[c]Partial gene sequences reported by Margis-Pinheiro et al. (2005).

[d]Novel sequence with verified biochemical activity matching that from the original reference given on the right.

[e]As identified in this study, OsKSL8 makes significant amounts of stemodene in addition to stemarene.

Table 2: OsKSL cloning primers.

OsKS1-F	*CACC*ATGAGGGACCAGCTCCAGACATTGGA
OsKS1-R	TCAATTGCCCGACAAAATAGAGCCATATGGAT
OsKSL2frag-F	CAGCTCGGCTCCTCGCCGGAGA
OsKSL2frag-R	ACATACCGGCTCTGCCTCCACTC
OsKSL3-F	*CACC*ATGTTTCAGTTAGAATTAGTGAACGTCGTC
OsKSL3-R	TCACGAAGCAGGAATGATATATATGGGTTC
OsKSL5-F	*CACC*ATGATACTTCCTATGAGTTCAGCATGCTT
OsKSL5-R	TCACAGCGTTCCCAAACCAGATGGAAG
OsKSL6-F	*CACC*ATGATGCTTCCTATGAGTTCAGCATGC
OsKSL6-R	TCACAGTGTTCCCGAATCAGATGGAG
OsKSL7-F	*CACC*ATGATGCTGCTAGGTTCCCCT
OsKSL7-R	CTACAATAATCTGAGTTGAAG
OsKSL8-3'RACE-F	*CACC*ATGATGCTGCTGAGTTCCTC
OsKSL8-3'frag-F	TCTGCTGTTGTCCCCTTGTTCCCCTGC
OsKSL8-3'frag-R	TTACTCTTGCAGGTGCAGTGGCTCCAGA
OsKSL10-F	*CACC*ATGGTAAGAAAACAGTTGCAGAGAG
OsKSL10-R	TCATGAGGACAACGTTTCTGG

Italicized *CACC* indicates introduced sequence required for directional topoisomerization mediated cloning.

Chapter IV. A modular approach for facile biosynthesis of labdane-related diterpenes

Reproduced with permission from Cyr, A., Wilderman, P. R., Determan, M., Peters, R. J. *J Am Chem Soc.,* **2007**, 129, 6684-6685.

Anthony Cyr, P. Ross Wilderman, Mara Determan, and Reuben J. Peters

ABSTRACT

Labdane-related diterpenoids are a large group of over 5,000 natural products whose biosynthesis typically proceeds through a labdadienyl/copalyl diphosphate (CPP) intermediate to a further cyclized and/or rearranged hydrocarbon diterpene en route to more elaborated compounds. Here we report a modular approach for facile biosynthesis of labdane-related diterpenes wherein base pGGxC vectors capable of introducing bacterial production of any one of the three common stereoisomers of CPP can be co-introduced with diterpene synthases that convert these CPP intermediates to specific diterpene hydrocarbon skeletal structures. The utility of this approach is demonstrated by individually engineering *E. coli* to produce any one of eight different diterpene skeletal structures, which collectively serve as precursors to literally thousands of distinct natural products.

MAIN TEXT

Labdane-related diterpenoids comprise a large group of over 5,000 known natural products defined as minimally containing the fused bicyclic hydrocarbon structure found in the labdane family of diterpenoids. This characteristic core structure results from the unusual biosynthetic origins of these compounds, which is uniquely initiated by a sequential pair of

terpene synthase catalyzed reactions. In particular, cyclization of the universal diterpenoid precursor (*E,E,E*)-geranylgeranyl diphosphate (GGPP, **1**) to a specific stereoisomer of labdadienyl/copalyl diphosphate (CPP, e.g. **2**-**4**) in a carbon-carbon double bond protonation-initiated reaction catalyzed by class II diterpene cyclases. This core bicyclic structure is then further cyclized and/or rearranged in diphosphate ionization initiated reactions catalyzed by more typical class I, but generally CPP stereospecific, labdane-related diterpene synthases, as in the related/derived structural families (e.g. kauranes, abietanes, and pimaranes).[1]

Included in the group of labdane-related diterpenoids are the gibberellin phytohormones required for normal growth and development in all higher plants, which has provided a ready reservoir of biosynthetic genes whose duplication has led to the widespread occurrence of such natural products throughout the plant kingdom. However, the vast majority of these compounds are secondary metabolites only found in a subset of species, in limited quantities, and often only under certain conditions. Thus, although several such compounds are utilized in industry, research, and medicine (e.g. resin acids and forskolin), similar uses for the bulk of this group of natural products have not been well explored. The use of recombinant microbial hosts for production of specific isoprenoid/terpenoid natural products has been recently reviewed.[2] To provide a foundation for broader investigations of labdane-related diterpenoid natural products we report here development of a modular approach that facilitates recombinant bacterial expression of various combinations of the sequentially acting class II and class I diterpene synthases, to produce hydrocarbon diterpene precursors to literally thousands of known natural products.

Because diterpene biosynthesis in plants occurs in plastids, recombinant microbial expression of the associated enzymes requires construction of pseudo-mature genes missing the N-terminal plastid targeting peptide sequence that interferes with proper protein folding.[3] This was recently accomplished for the *ent*-CPP synthase from *Arabidopsis thaliana*

(AtCPS),[4] providing a guide for construction of pseudo-mature versions of other strictly class II diterpene cyclases and, thus, enabling the studies reported here.

While all organisms produce terpenoids/isoprenoids, *E. coli* does not typically produce GGPP, hence it was necessary to introduce a GGPP synthase (GGPS), which was done through the use of a previously reported recombinant pseudo-mature GGPS from grand fir (*Abies grandis*), rAgGGPS.[5] To minimize the number of plasmids required for these metabolic engineering efforts, rAgGGPS was introduced into the dual gene expression vector pACYCDuet, creating a base pGG plasmid (see Supporting Information). For initial proof-of-concept studies a previously reported pseudo-mature version of the bifunctional class II/I diterpene cyclase from grand fir, abietadiene synthase (rAgAS)[6] was added to pGG to create pGGAS. Transformation with pGGAS yielded recombinant bacteria that produce the expected abietadiene (**5**) diterpene upon induction of the encoded enzymatic genes (Scheme 1).

From this initial success we were encouraged to attempt the expression of more typical separate class II and class I labdane-related diterpene synthases. In particular, because we have recently reported characterization of a pseudo-mature version of AtCPS, this gene was incorporated into pGG to create pGG*e*C, which should enable bacterial production of *ent*-CPP (**2**). We also have reported functional recombinant expression of tagged (e.g. with glutathione-*S*-transferase, GST) versions of class I labdane-related diterpene synthases,[7] including that for the *ent*-CPP specific kaurene synthase (KS) from *Arabidopsis thaliana* (AtKS),[8] using pET based expression vectors (e.g. pDEST), which can be co-transformed with pACYCDuet and, hence, the derived pGG and pGG*e*C. Co-transformation of pGG*e*C with a pDEST15/AtKS vector encoding a GST-AtKS fusion protein yielded recombinant bacteria that produce the expected *ent*-kaur-16-ene (**6**) (Figure 1).

More detailed analysis of this kaurene (**6**) producing strain was carried out to optimize diterpene production. Somewhat surprisingly, ~90% of the hydrocarbon product

was secreted to the media, which had not been previously reported.[9] Similar secretion was observed with all the labdane-related diterpenes we have tested. Accordingly, it is possible to trap these hydrocarbon compounds on hydrophobic (Diaion HP-20) beads under mixed phase liquid-solid (i.e. media-beads) growth conditions,[10] and easily isolate the produced diterpenes (e.g. Figure 1). With the optimized conditions it was possible to obtain ~100 μg of kaurene (**6**)/L mixed phase bacterial culture, similar to previously reported terpenoid yield from *E. coli* that has not been engineered for increased isoprenoid precursor supply.[11]

The vast majority of labdane-related diterpenoid natural products are derived from three of the four possible stereoisomers of CPP, specifically, *ent*- (**2**), *syn*- (**3**), or normal (**4**) CPP.[1] The pGG*e*C construct provides access to *ent*-CPP derived diterpenes through co-expression of class I labdane-related diterpene synthases. To provide similar access to *syn*-CPP (**3**) derived diterpenes we utilized the *syn*-CPP diterpene cyclase from rice (*Oryza sativa*) we have previously identified (OsCPS4)[12] and, based on our work with rAtCPS,[4] constructed a pseudo-mature version suitable for recombinant bacterial expression (rOsCPS4, missing amino acids 2-70). Insertion of rOsCPS4 into pGG created pGG*s*C. To provide access to normal CPP (**4**) derived diterpenes, we returned to abietadiene synthase, which produces normal CPP (**4**) as an intermediate, and for which a D621A mutant (rAgAS:D621A) that no longer has class I activity [i.e. only produces **4** from GGPP (**1**)] has been reported.[13] Thus, rAgAS:D621A was inserted into pGG to create pGGnC.

To demonstrate the ability of these three pGGxC constructs to provide access to various stereoisomers of CPP for the production of a range of CPP derived diterpenes, pGG*e*C and pGG*s*C were individually co-transformed with previously described class I enzymes specific for the corresponding stereoisomer of CPP,[7] while pGGnC was co-transformed with a D404A mutant of rAgAS (rAgAS:D404A) that only catalyzes the production of abietadiene (**5**) from **4** (i.e. only has class I activity).[13] In each case this resulted in recombinant bacteria that produce the expected labdane-related diterpene (Table

1). Eight different hydrocarbon skeletal structures, with one as distinct double bond isomers, were produced, proving the utility of our modular approach for combinatorial biosynthesis of labdane-related diterpenes (Scheme 1). Notably, more than 1,000 natural products are derived from kaurene, with almost as many being derived from abietadiene, as well as ~500 from various pimaradienes.[14] Thus, the diterpenes produced in this study are precursors to literally thousands of known compounds, and these results establish a foundation for biosynthetic production of elaborated labdane-related diterpenoid natural products. Both in providing substrates for characterization of downstream biosynthetic steps and for further extension of metabolic engineering efforts. In addition, the three pGGxC constructs reported here will be useful for characterization of novel class I labdane-related diterpene synthases. Finally, we expect that a similar approach will provide access to labdane-related diterpenes and diterpenoid natural products derived from other class II intermediates such as clerodadienyl diphosphate (i.e. by incorporating class II diterpene cyclases that produce such compounds[15] into pGG), which will further broaden the reach of the modular approach to labdane-related diterpene biosynthesis demonstrated here.

ACKNOWLEDGEMENTS

We thank Professor Robert M. Coates (Univ. Illinois) for authentic standards, Professor Rodney B. Croteau (Washington State Univ.) for the rAgGGPS and rAgAS clones, and Luke Lowry, Dana Morrone, and Meimei Xu for technical support. This work was funded by grants from the NSF (MCB-0416948) and NIH (GM076324) to R.J.P.

† Current address: University of Iowa, Medical Scientist Training Program

REFERENCES

1. MacMillan, J.; Beale, M. H., Diterpene biosynthesis. In *Isoprenoids Including Carotenoids and Steroids*, 1st ed.; Cane, D. E., Ed. Elsevier Science Ltd.: Oxford, 1999; Vol. 2, pp 217-243.

2. Chang, C. Y. M.; Keasling, J. D., *Nat. Chem. Biol.* **2006,** 2, 674-681.

3. Croteau, R.; Kutchan, T. M.; Lewis, N. G., Natural products (secondary metabolites). In *Biochemistry & Molecular Biology of Plants*, Buchanan, B.; Gruissem, W.; Jones, R., Eds. Am. Soc. Plant Biologists: Rockville, MD, USA, 2000; pp 1250-1318.

4. Prisic, S.; Peters, R. J., *Plant Physiol.* **2007,** 144, 445-454.

5. Burke, C.; Croteau, R., *J. Biol. Chem.* **2002,** 277, 3141-3149.

6. Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H.-J.; Coates, R. M.; Croteau, R. B., *Biochemistry* **2000,** 39, 15592-15602.

7. (a) Wilderman, P. R.; Xu, M.; Jin, Y.; Coates, R. M.; Peters, R. J., *Plant Physiol.* **2004,** 135, 2098-2105. (b) Morrone, D.; Jin, Y.; Xu, M.; Choi, S.-Y.; Coates, R. M.; Peters, R. J., *Arch. Biochem. Biophys.* **2006,** 448, 133-140. (c) Xu, M.; Wilderman, P. R.; Morrone, D.; Xu, J.; Roy, A.; Margis-Pinheiro, M.; Upadhyaya, N.; Coates, R. M.; Peters, R. J., *Phytochemistry* **2007,** 68, 312-326.

8. Roy, A.; Roberts, F. G.; Wilderman, P. R.; Peters, R. J.; Coates, R. M., *J. Am. Chem. Soc.* **2007,** in press.

9. Reiling, K. K.; Yoshikuni, Y.; Martin, V. J. J.; Newman, J.; Bohlmann, J.; Keasling, J. D., *Biotechnol. Bioeng.* **2004,** 87, 200-212.

10. Jackson, B. E.; Hart-Wells, E. A.; Matsuda, S. P. T., *Org. Lett.* **2003,** 5, 1629-1632.

11. Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D., *Nature Biotech.*, **2003,** 21, 796-802.

12. Xu, M.; Hillwig, M. L.; Prisic, S.; Coates, R. M.; Peters, R. J., *Plant J.* **2004,** 39, 309-318.

13. Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B., *J. Am. Chem. Soc.* **2001,** 123, 8974-8978.

14. Buckingham, J., *Dictionary of Natural Products (on-line web edition)*. Chapman & Hall/CRC Press: 2002.

15. Hamano, Y.; Kuzuyama, Y.; Itoh, N.; Furihata, K.; Seto, H.; Dairi, T., *J. Biol. Chem.* **2002,** 277, 37098-37104.

SUPPORTING INFORMATION

General.

Unless otherwise noted all chemicals were purchased from Fisher Scientific (Loughborough, Leicestershire, UK) and molecular biology reagents from Invitrogen (Carlsbad, CA, USA).

Vector construction.

pGG

The pseudo-mature geranylgeranyl diphosphate synthase construct rAgGGPS[1] was PCR amplified using primers to introduce a 5' NdeI restriction site that overlapping with the initiating ATG codon and a 3' XhoI site immediately following the stop codon. The rAgGGPS PCR product and pACYCDuet (Novagen, San Diego, CA) were digested first with NdeI and then XhoI. The resulting large fragments were gel purified and then ligated together to create pGG. The inserted rAgGGPS, which falls within the second pACYCDuet multiple cloning site (MCS2), was verified by complete sequencing.

pGGAS

The pseudo-mature abietadiene synthase rAgAS[2] was inserted into MCS1 from pACYCDuet in pGG. A 5' CAT overhang at the 5' end of rAgAS, where the AT is part of the

initiating ATG codon, was created by sticky-end PCR[3] and a 3' NotI site immediately following the stop codon also introduced. After digestion of the resulting sticky-end PCR product with NotI, this rAgAS construct was ligated into NcoI/NotI digested pGG to create pGGAS, and the presence of the rAgAS insert verified by sequencing the 5' and 3' ends.

pGGeC

The pseduo-mature *ent*-copalyl diphosphate synthase rAtCPS[4] was similarly inserted into MCS1 from pACYCDuet in pGG (i.e. using sticky-end PCR to create a 5' CAT overhang and 3' NotI site for digestion and ligation into NcoI/NotI digested pGG) to create pGG*e*C. The inserted rAtCPS was also verified by complete sequencing.

pGGsC

Three different pseduo-mature constructs, missing 70, 82, or 86 amino acid (aa) residues, of the *syn*-copalyl diphosphate synthase OsCPS4[5] were created (Δ70, Δ82, and Δ86, respectively). Preliminary analysis suggested that OsCPS4Δ70, which parallels rAtCPS in being truncated 32 aa residues upstream of a conserved SAYDT motif, exhibited the optimal combination of expression level and biochemical activity. Both NcoI sites (at nucleotides 903 and 1205) of OsCPS4Δ70 (rOsCPS4) were removed by synonymous site-directed mutagenesis using PCR amplification with overlapping mutagenic primers. The resulting rOsCPS4(-NcoI) construct was verified by complete sequencing, then PCR amplified with primers that introduced a 5' NcoI site overlapping with the initiating ATG codon and a 3' NotI site immediately after the stop codon. This PCR product, along with pGG, was digested with NcoI and then NotI, and the resulting large fragments gel purified and then ligated together to create pGG*s*C. The inserted rOsCPS4 was verified by complete sequencing.

pGGnC

Both NcoI sites of rAgAS (at nucleotides 972 and 1630) were removed by synonymous site-directed mutagenesis using PCR amplification with overlapping mutagenic primers. The resulting rAgAS(-NcoI) construct was verified by complete sequencing and the D621A mutant[6] (re)created by overlapping PCR site-directed mutagenesis. This rAgAS:D621A construct was PCR amplified with primers that introduced a 5' NcoI site overlapping with the initiating ATG codon and a 3' BamHI site immediately after the stop codon. This PCR product, along with pGG, was digested with NcoI and then BamHI, and the resulting large fragments gel purified and then ligated together to create pGGnC. The inserted rAgAS:D621A was verified by complete sequencing.

Recombinant strains.

Analysis

Gas chromatography-mass spectrometry (GC-MS) was performed on organic extracts using an HP1-MS column on an Agilent (Palo Alto, CA) 6890N GC instrument with 5973N mass selective detector in electron-ionization mode (70eV) located in the W.M. Keck Metabolomics Research laboratory at Iowa State University, much as previously described.[5] Briefly, 5 µL of sample was injected at 40°C in splitless mode, the oven temperature held at 40°C for 3 min., then raised at 20°C/min. to 300°C, and held there for 3 min. MS data was collected from 50 to 500 *m/z* during the temperature ramp and final hold. The biosynthetically produced diterpenes were identified by comparison of retention time and mass spectra to authentic samples (see Supporting Figure). The amount of diterpene produced was determined by comparison of the biosynthetic organic extracts to a standard curve constructed with the use of known quantities of cembrene (Acros Organics, Geel, Belgium).

Transformation and expression

Previous work has established that the OverExpress C41 strain of *E. coli* (Lucigen, Middleton, WI) is well suited to T7 promoter based expression of plant derived labdane-related diterpene synthases.[4] Thus, all work in this study has been carried out with the use of this C41 strain. Chemically competent C41 cells (10 μL) were typically co-transformed with 1 μL of each plasmid and grown with dual antibiotic (34 μg/mL chloramphenicol and 50 μg/mL carbenicillin) selection on NZY media. Expression of the encoded genes was induced by the addition of IPTG.

Optimization

A number of factors were examined for their effect on diterpene production in the utilized shake flask culture growths. In particular, with the *ent*-kaur-16-ene (**6**) producing C41 strain transformed with pGG*e*C + pDEST15/AtKS. First, while all cultures were grown to mid-log phase (0.4-0.6 A_{600}) at 37°C, the effect of varying temperature from 15-30°C for expression of the introduced metabolic pathway was examined, and 20°C found to be optimum. Second, the amount of IPTG inducer, as a proxy for enzymatic expression level, was varied between 0.05 - 1 mM, and 0.5 mM found to be optimal. Thus, all cultures were grown to mid-log phase at 37°C, then shifted to 20°C for 1 hour prior to induction of the introduced metabolic pathway with 0.5 mM IPTG. Finally, the addition of glycerol to a final concentration of 5 g/L was found to also increase diterpene yield.

Secretion of diterpenes enables ready isolation from mixed phase culture growths

Expecting the produced hydrocarbon compounds to partition into the membrane, our initial analyses of diterpene production solely focused on organic solvent extraction of cell pellets from shake flask growths. However, more detailed analysis of the *ent*-kaur-16-ene (**6**)

producing C41 strain transformed with pGG*e*C + pDEST15/AtKS revealed that ~90% of the diterpene resides in the spent culture media (i.e. the liquid remaining after the cells were removed by centrifugal pelleting). The spent media also provided a cleaner source of diterpene than cell pellets. In addition, the use of mixed phase cultures, specifically liquid media growth cultures containing adsorbent resin (i.e. liquid and solid phases), for harvesting a similarly secreted sesquiterpenoid has been previously reported.[7] We also found that inclusion of hydrophobic Diaion HP-20 beads (Supleco, Bellefonte, PA) enabled ready isolation of the biosynthetically produced diterpenes. Upon analysis of such mixed phase cultures with 1-5% (wt/vol) Diaion HP-20 beads, growths with 2% Diaion HP-20 beads were found to be optimal, increasing the observed yield ~2-fold from that obtained with liquid only cultures (presumably by preventing the loss of diterpene to volatilization, e.g. kaurene is emitted by several plant species,[8] and similar loss to volatilization has been reported for *E. coli* engineered to produce the sesquiterpene amorpha-4,11-diene[9]). The Diaion HP-20 beads from a typical 50 mL mixed phase culture growth were removed by filtration through a #100 sieve (Hogentogler & Co, Columbia, MD), washed in 25 mL dH_2O, again isolated by filtration, and then washed in 25 mL EtOH, followed by transfer to a glass wool plugged pasteur pipet. The isolated beads were then further washed with EtOH until all pigmentation is removed, i.e. the eluant is clear (10-15 column volumes). The beads were then dried by the application of a gentle flow of nitrogen gas prior to elution with 6 mL hexanes. A ~1 mL silica gel column with ~0.1 mL anhydrous magnesium sulfate overlay pre-washed with ~2 mL hexanes was prepared, and the organic extract passed over this column, which was then further washed with an additional ~2 mL hexanes. The pooled extract and wash was then dried under nitrogen gas and resuspended in 1 mL of hexanes for GC-MS analysis. Typical yields from this protocol were 5-100 μg/L of mixed phase culture growth (see Table S1), with the diterpene generally representing the most abundant extracted compound (e.g. Figure 1). Analysis of strains carrying pGG*e*C + either a mutant or wild-type rAtKS indicate that

product yield corresponds to KS activity measured in vitro,[10] suggesting that the variation in yield observed here may largely reflect underlying changes in total activity, due to differences in catalytic rate and/or expression level, of the various class I diterpene synthases utilized in these studies. Notably, the yields at the higher end of the range observed here are similar to that previously reported for *E. coli* engineered to produce the sesquiterpene amorpha-4,11-diene prior to any further optimization of isoprenoid precursor supply (ca ~100 μg/L in similar shake-flask production runs), although a >100-fold in yield was observed upon such optimization.[11] Thus, future work will be focused on similarly increasing isoprenoid precursor supply within the context of the modular approach to labdane-related diterpene biosynthesis demonstrated here.

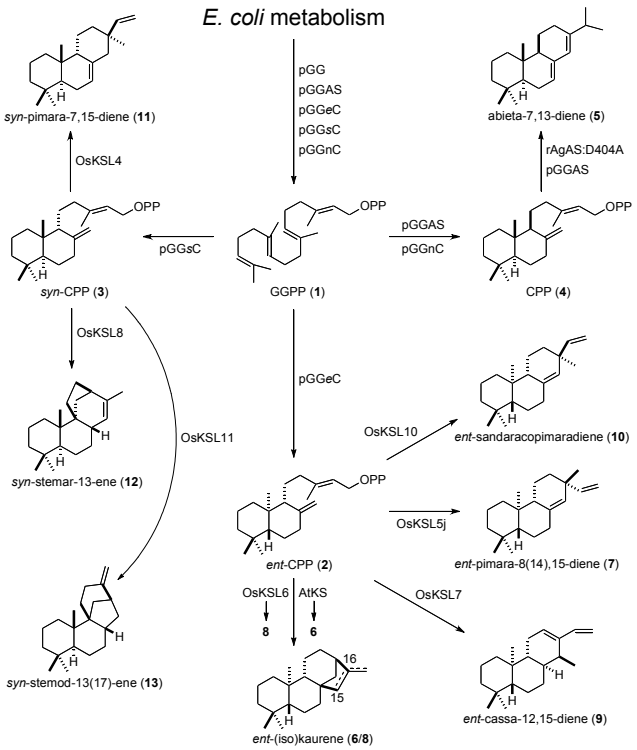
REFERENCES FOR SUPPORTING INFORMATION

1. Burke, C.; Croteau, R. *J. Biol. Chem.* **2002**, *277*, 3141-3149.
2. Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H.-J.; Coates, R. M.; Croteau, R. B. *Biochemistry* **2000**, *39*, 15592-15602.
3. Zeng, G. *Biotechniques* **1998**, *25*, 206-208.
4. Prisic, S.; Peters, R. J. *Plant Physiol.* **2007**, *144*, 445-454.
5. Xu, M.; Hillwig, M. L.; Prisic, S.; Coates, R. M.; Peters, R. J. *Plant J.* **2004**, *39*, 309-318.
6. Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B. *J. Am. Chem. Soc.* **2001**, *123*, 8974-8978.
7. Jackson, B. E.; Hart-Wells, E. A.; Matsuda, S. P. T. *Org. Lett.* **2003**, *5*, 1629-1632.
8. Otsuka, M.; Kenmoku, H.; Ogawa, M.; Okada, K.; Mitsuhashi, W.; Sassa, T.; Kamiya, Y.; Toyomasu, T.; Yamaguchi, S. *Plant Cell Physiol.* **2004**, *45*, 1129-1138.
9. Newman, J. D.; Marshall, J.; Chang, M.; Nowroozi, R.; Paradise, E.; Pitera, D.; Newman, K. L.; Keasling, J. D. *Biotechnol. Bioeng.* **2006**, *95*, 684-691.

10. Xu, M.; Wilderman, P. R.; Peters, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 7397-7401.

11. Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. *Nat. Biotechnol.* **2003**, *21*, 796-802.

Scheme 1. Modular approach to labdane-related diterpene biosynthesis.



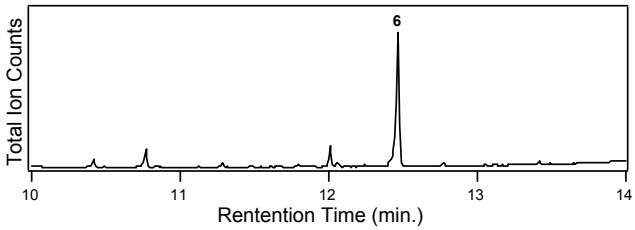


Figure 1. Production of *ent*-kaur-16-ene (**6**) by *E. coli* transformed with pGG*e*C and pDEST15/rAtKS. Total ion chromatogram from GC-MS analysis of organic solvent eluant of Diaion HP-20 beads from a mixed phase culture (see Supporting Information).

Table 1. Recombinant bacteria characterized in metabolic engineering study.

Plasmid(s)	Product
pGGAS	abieta-7,13-diene (**5**)
pGG*e*C + pDEST15/rAtKS	*ent*-kaur-16-ene (**6**)
pGG*e*C + pDEST15/OsKSL5j	*ent*-pimara-8(14),15-diene (**7**)
pGG*e*C + pTH1/OsKSL6	*ent*-isokaur-15-ene (**8**)
pGG*e*C + pDEST15/OsKSL7	*ent*-cassa-12,15-diene (**9**)
pGG*e*C + pDEST15/OsKSL10	*ent*-sandaracopimaradiene (**10**)
pGG*s*C + pDEST14/OsKSL4	*syn*-pimara-7,15-diene (**11**)
pGG*s*C + pTH8/OsKSL8	*syn*-stemar-13-ene (**12**)
pGG*s*C + pDEST15/OsKSL11	*syn*-stemod-13(17)-ene (**13**)
PGGnC + pDEST14/rAgAS:D404A	abieta-7,13-diene (**5**)

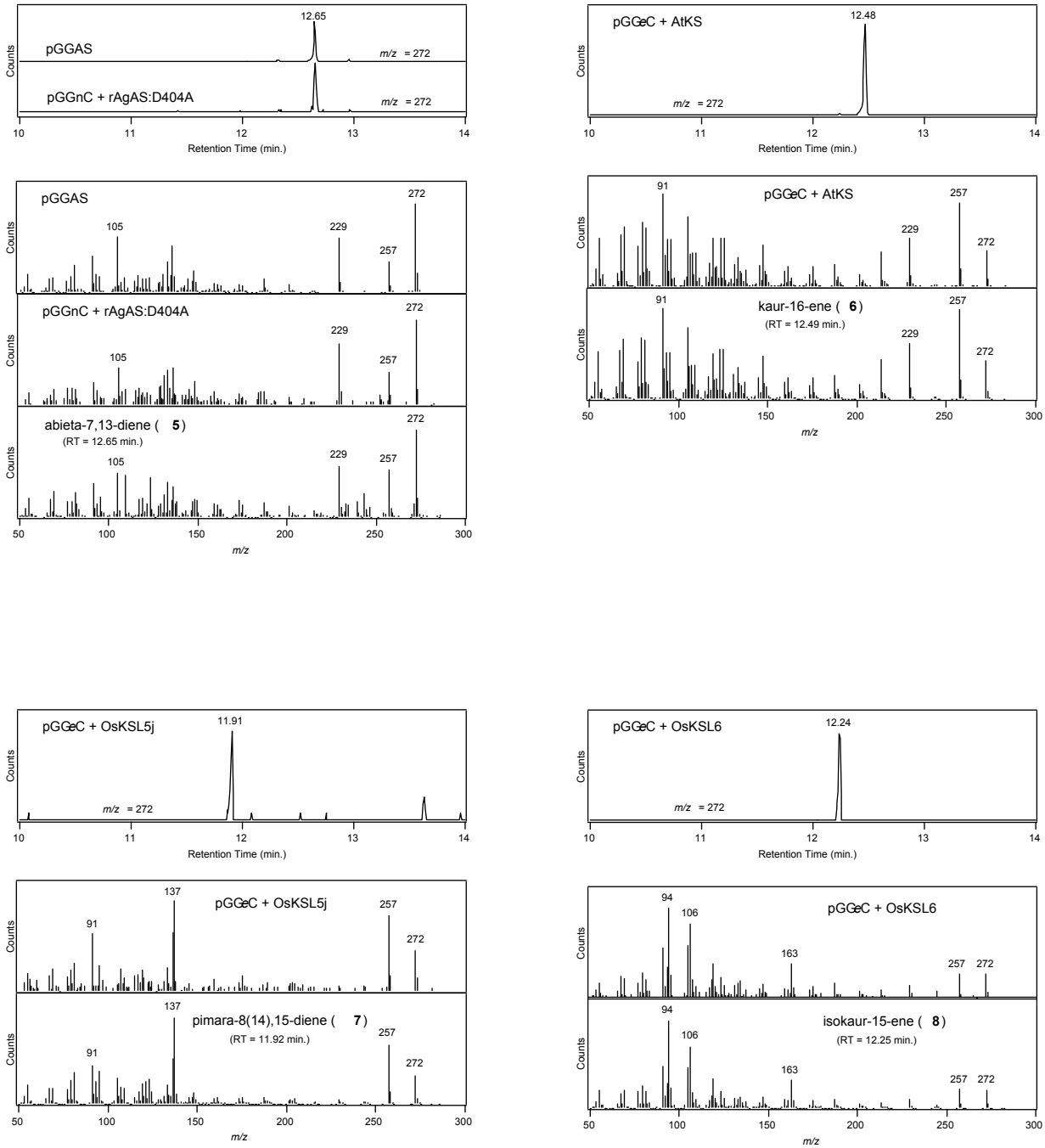
E. coli containing the indicated plasmid(s) produce the indicated labdane-related diterpene, as identified by GC-MS based comparison to authentic standards, with production levels ranging from 5-100 μg diterpene/L mixed phase culture (see Supporting Information).

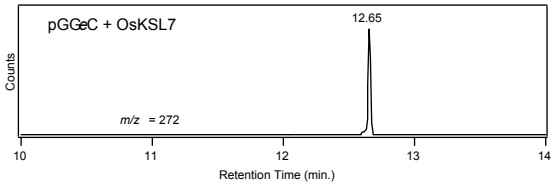
Table S1: Characterization of engineered *E. coli*.

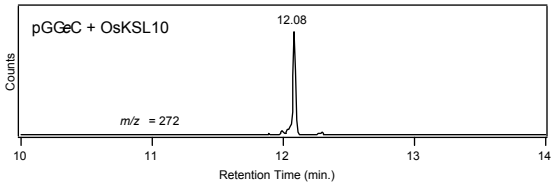
Plasmid(s)	Product	Yield (µg/L culture)*
pGGAS	abieta-7,13-diene (**5**)	50
pGG*e*C + pDEST15/rAtKS	*ent*-kaur-16-ene (**6**)	95
pGG*e*C + pDEST15/OsKSL5j	*ent*-pimara-8(14),15-diene (**7**)	11
pGG*e*C + pTH1/OsKSL6	*ent*-isokaur-15-ene (**8**)	81
pGG*e*C + pDEST15/OsKSL7	*ent*-cassa-12,15-diene (**9**)	67
pGG*e*C + pDEST15/OsKSL10	*ent*-sandaracopimaradiene (**10**)	90
pGG*s*C + pDEST15/OsKSL4	*syn*-pimara-7,15-diene (**11**)	32
pGG*s*C + pTH8/OsKSL8	*syn*-stemar-13-ene (**12**)	20
pGG*s*C + pDEST15/OsKSL11	*syn*-stemod-13(17)-ene (**13**)	70
pGGnC + pDEST14/rAgAS:D404A	abieta-7,13-diene (**5**)	5

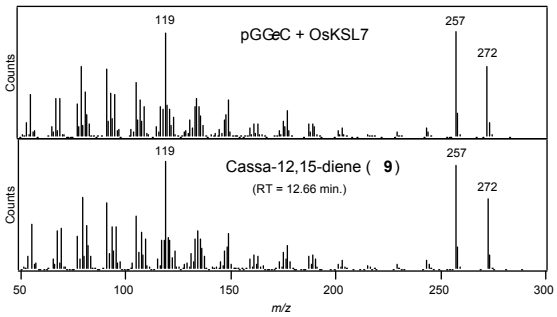
*Yield from repeated runs is typically within 20% of that reported here.

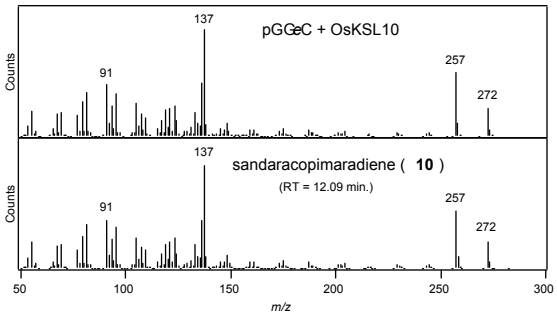
Figure S1: Selective ion chromatograms from GC-MS analysis of metabolically engineered *E. coli*, with the retention time (RT) of the biosynthetically produced diterpene indicated. Also shown are the mass spectra for the biosynthetically produced diterpenes and authentic reference compounds (with their RT indicated in parentheses).

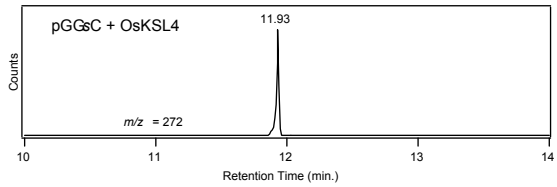


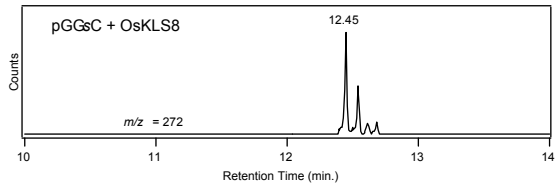

pGGeC + OsKSL7
12.65
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)

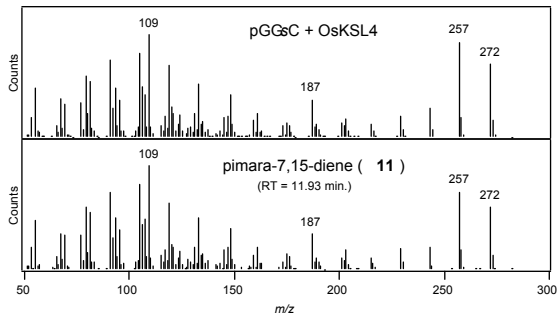

pGGeC + OsKSL10
12.08
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)

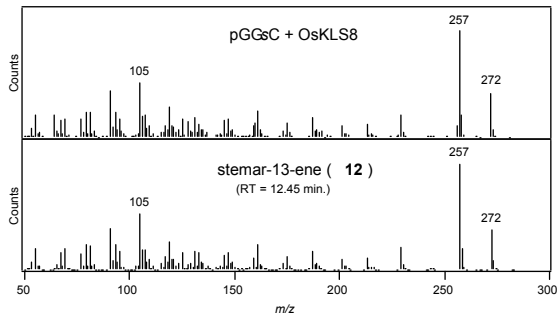

pGGeC + OsKSL7
119
257
272
Counts
Cassa-12,15-diene (9)
(RT = 12.66 min.)
119
257
272
Counts
50
100
150
200
250
300
m/z

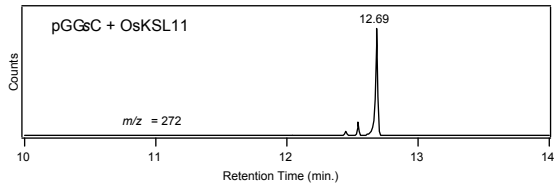

pGGeC + OsKSL10
137
91
257
272
Counts
sandaracopimaradiene (10)
(RT = 12.09 min.)
137
91
257
272
Counts
50
100
150
200
250
300
m/z

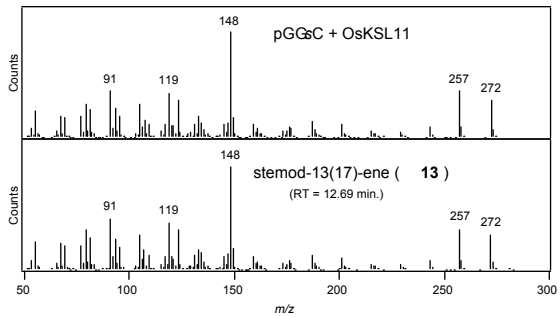

pGGsC + OsKSL4
11.93
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)


pGGsC + OsKLS8
12.45
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)


pGGsC + OsKSL4
109
257
272
187
Counts
pimara-7,15-diene (11)
(RT = 11.93 min.)
109
257
272
187
Counts
50
100
150
200
250
300
m/z


pGGsC + OsKLS8
257
105
272
Counts
stemar-13-ene (12)
(RT = 12.45 min.)
257
105
272
Counts
50
100
150
200
250
300
m/z


pGGsC + OsKSL11
12.69
Counts
m/z = 272
10
11
12
13
14
Retention Time (min.)


pGGsC + OsKSL11
148
91
119
257
272
Counts
stemod-13(17)-ene (13)
(RT = 12.69 min.)
148
91
119
257
272
Counts
50
100
150
200
250
300
m/z

Chapter V. Following evolution's lead to a single residue switch for diterpene synthase product outcome.

Reproduced with permission from Xu, M., Wilderman, P. R., Peters, R. J. (2007) *Proceedings of the National Academy of Sciences of the United States of America* **104,** 7397-7401.

Meimei Xu, P. Ross Wilderman, and Reuben J. Peters

ABSTRACT

There have been few insights into the biochemical origins of natural products biosynthesis from primary metabolism. Of particular interest are terpene synthases, which often mediate the committed step in particular biosynthetic pathways, so that alteration of their product outcome is a key step in derivation of novel natural products. These enzymes also catalyze complex reactions of significant mechanistic interest. Following an evolutionary lead from two recently diverged functionally distinct rice subspecies associated diterpene synthase orthologs we have identified a single residue that can act to switch product outcome. Specifically, mutation of a conserved isoleucine to threonine that acts to convert not only the originally targeted isokaurene synthase into a specific pimaradiene synthase but also much more broadly, including conversion of the *ent*-kaurene synthases found in all higher plants for gibberellin phytohormone biosynthesis to the production of pimaradiene. This surprisingly facile switch for diterpene synthase catalytic specificity indicates the ease with which primary (gibberellin) metabolism can be subverted to secondary biosynthesis, and may underlie the widespread occurrence of pimaradiene derived natural products. In addition, because this isoleucine is required for the mechanistically more complex cyclization to tetracyclic kaurene, while substitution with threonine 'short circuits'

this mechanism to produce the 'simpler' tricyclic pimaradiene, our results have novel implications regarding the means by which terpene synthases specify product outcome.

INTRODUCTION

The evolution of secondary metabolism presumably originates via changes in the catalytic specificity of enzymes recruited from primary metabolism. Although a recent report demonstrates that small numbers of changes in a sesquiterpene synthase can dramatically shift product outcome (i.e. plasticity), the target enzyme was already involved in secondary metabolism and the parent wild type enzyme was quite promiscuous (i.e. produced many different products) (1). Thus, how readily the typically specific enzymes involved in primary metabolism can be subverted into secondary metabolism remains a matter of conjecture.

Terpene synthases carry out complex electrophilic cyclization/rearrangement reactions, creating diverse hydrocarbon skeletal structures from simpler isoprenoid precursors, which often represents the committed step in particular biosynthetic pathways. Hence, altering the function of these enzymes is expected to represent a key step in the evolution of novel secondary metabolism. In addition, terpene synthases have attracted a great deal of interest due to their complex reaction mechanisms and wide variety of resulting products; in particular how these enzymes specify product outcome (2), most recently with reports demonstrating that specificity can be dramatically shifted by changes in a small number of amino acid residues (1, 3, 4).

Kaurene synthase (KS) catalyzes the cyclization of *ent*-copalyl diphosphate (*ent*-CPP, **1**) to *ent*-kaur-16-ene (**2**) through a multiple step reaction mechanism (Scheme 1) (5). This enzyme is found in all higher plants, as kaurene is an intermediate en route to the diterpenoid

gibberellin phytohormones required for normal growth and development (6). Hence, KS participates in primary metabolism.

The rice (*Oryza sativa*) genome contains an expanded family of KS like (KSL) genes with differing function (7). In the course of our biochemical characterization of the rice KSL gene family (OsKSL) we found that one such enzyme from subspecies *indica* (OsKSL5i) cyclized *ent*-CPP (**1**) specifically (>90%) to *ent*-isokaur-15-ene (**3**) (8). However, another group had previously reported that the orthologous OsKSL5j from subspecies *japonica* rice specifically produced *ent*-pimara-8(14),15-diene (**4**), representing deprotonation of the putative pimaren-8-yl carbocation (**5**) intermediate in cyclization of *ent*-CPP (**1**) to *ent*-(iso)kaurene (Scheme 1), instead (9). Here we demonstrate that alternation of a single residue is sufficient for the change in product outcome between these two orthologous diterpene synthases, and extension of these results to disparate kaurene synthases involved in gibberellin biosynthesis, with significant implications for not only how terpene synthases control their product outcome but the evolution of secondary metabolism as well.

RESULTS

The functionally distinct rice subspecies associated orthologs OsKSL5i and OsKSL5j are 98% identical at the amino acid (aa) level and, from modeled structures, there are only three differences in the active site (8). To determine which, if any, was responsible for the observed change in function, each of these three divergent active site residues in the *ent*-isokaurene synthase OsKSL5i was mutated to the corresponding amino acid found in the *ent*-pimaradiene synthase OsKSL5j. The product profile of the resulting OsKSL5i:V661L, OsKSL5i:I664T, and OsKSL5i:I718V mutants was characterized by GC-MS analysis and

comparison of the enzymatic product to authentic standards. While OsKSL5i:V661L and OsKSL5i:I718V continued to specifically produce *ent*-isokaur-15-ene (**3**), OsKSL5i:I664T specifically produced *ent*-pimara-8(14),15-diene (**4**) (Figure 1). Thus, this single isoleucine to threonine change is sufficient to convert the specific isokaurene synthase OsKSL5i into a specific pimaradiene synthase.

Rice contains another *ent*-isokaurene synthase (OsKSL6) that shares 89% aa identity with OsKSL5i (8, 9), and the isoleucine residue identified above is conserved in OsKSL6 (also at position 664). To determine if the same change would similarly convert OsKSL6 into a *ent*-pimaradiene synthase the analogous OsKSL6:I664T mutant was constructed and its product profile characterized by GC-MS. Notably, this isoleucine to threonine change is also sufficient to convert OsKSL6 from a specific *ent*-isokaurene to specific *ent*-pimaradiene synthase (Figure 2A).

Although kaurene synthases are much less well conserved with OsKSL5i (41-52% aa identity), a corresponding isoleucine residue is further found in all the known KS (Figure 3). The KS most similar to OsKSL5i is the rice paralog OsKS1, sharing 52% aa identity (8). To ascertain if this isoleucine residue plays a similar role in KS as it does in OsKSL5i and OsKSL6 the analogous OsKS1:I602T mutant was constructed and characterized. Strikingly, this same isoleucine to threonine change also switches OsKS1 from a specific kaurene synthase to a pimaradiene synthase, albeit one that still produces some (~30% by GC-FID analysis) of *ent*-kaur-16-ene (**2**) (Figure 2B). In addition, while the KS from the dicot *Arabidopsis thaliana* (AtKS) is the most distantly related, sharing only 41% aa identity with OsKSL5i, the corresponding AtKS:I638T mutant also produces largely (~80%) *ent*-pimara-8(14),15-diene (**4**) along with smaller amounts (~20%) of *ent*-kaur-16-ene (**2**) (Figure 2C).

Thus, this single isoleucine to threonine change, requiring substitution of only a single nucleotide, is sufficient to essentially switch the product outcome of a wide range of (iso)kaurene synthases to the production of pimaradiene (**4**).

To further probe the role of the residue at this position in determining product outcome, particularly for secondary (tetra)cyclization within the initially formed tricyclic pimaren-8-yl$^+$ (**5**) intermediate (Scheme 1), OsKSL5j was cloned and the converse threonine to isoleucine mutant constructed and characterized. Intriguingly, OsKSL5j:T664I no longer produces *ent*-pimara-8(14),15-diene, but nor does it exclusively produce isokaur-15-ene (**3**). Instead, this mutant produces a mixture of tetracyclic diterpenes (Figure 3), exhibiting a 50:37:13 ratio of isokaur-15-ene (**3**) to atiser-16-ene (**6**) to kaur-16-ene (**2**). While OsKSL5i does produce small amounts of atiser-16-ene (**6**) and kaur-16-ene (**2**) (~10% altogether), the product profile of the OsKSL5j:T664I mutant demonstrates decreased product specificity resulting from one or more of the other differences between OsKSL5i and OsKSL5j. Regardless, the specific production of tetracyclic diterpenes by OsKSL5j:T664I indicates the importance of an isoleucine at this position for secondary cyclization of the pimaren-8-yl$^+$ intermediate (**5**) to a beyeran-16-yl$^+$ intermediate (Scheme 1).

In order for the observed single residue switch of diterpene synthase product outcome to be relevant in biological settings and play a role in biochemical evolution the effect of such a change on enzymatic activity must be minimal, such that effective metabolic flux towards the new product can be observed. To ascertain the effect of the reported mutations each of the wild type and mutant diterpene synthases was co-expressed in recombinant *E. coli* with a GGPP and *ent*-CPP synthase. In every case production of the expected diterpene(s) was

readily observed, demonstrating that these mutants are biologically functional, mediating observable metabolic flux at least in this heterologous setting. In addition, it has been reported that terpene biosynthesis in metabolically engineered *E. coli* is limited, at least in part, by terpene synthase activity (10), suggesting that terpene production levels can serve as a proxy for relative enzymatic activity. Quantitative product analysis (by GC-FID analysis) demonstrated that the isoleucine to threonine mutants reported above reduced the amount of diterpene(s) produced by 3 to 11-fold, indicating that these mutations had small but observable deleterious effects on catalytic activity. Further, the OsKSL5j:T664I mutant actually increased diterpene production ~15-fold over that mediated by wild type OsKSL5j, also indicating that the change from isoleucine to threonine at this position reduces catalytic activity. Finally, consistent with our hypothesis that the final amount of diterpene from these metabolically engineered *E. coli* is proportional to enzymatic activity, kinetic analysis of AtKS (apparent k_{cat} = 0.06±0.04 s^{-1} and K_M = 0.3±0.1 μM) and the AtKS:I638T mutant (apparent k_{cat} = 0.02±0.01 s^{-1} and K_M = 0.4±0.2 μM) demonstrated a ~4-fold decrease in catalytic efficiency [(2 versus 0.5) × 10^5 M^{-1} s^{-1}], similar to the ~3-fold reduction suggested by quantification of the amount of diterpene produced by metabolically engineered *E. coli*.

DISCUSSION

Given the complex catalytic mechanisms and wide variety of resulting products arising from terpene synthases, as well as their critical role in determining metabolic fate, how these enzymes specify product outcome has been a subject of great biochemical interest (2). Recent results indicate that small numbers of amino acid substitutions are sufficient to

dramatically alter the product profile of sesquiterpene synthases (1, 3, 4). However, there have been no previous reports targeting the catalytic specificity of diterpene synthases. In addition, previous reports have largely focused on inter-conversion of pairs of terpene synthases sharing significant similarity (>70% aa identity), although one recent study more generally investigated the plasticity of terpene synthase activity (1). Here, following a lead from two recently diverged rice subspecies associated diterpene synthase orthologs, a specific isoleucine to threonine change was found to be sufficient to essentially switch the product outcome of not only the originally targeted OsKSL5i *ent*-isokaurene synthase ortholog and a very similar OsKSL6 paralog, but also kaurene synthases exhibiting much less similarity (41-52% aa identity). In addition, OsKS1 and AtKS are among the most disparate pair of kaurene synthases, sharing only 47% aa identity, which is close to the low end of the 46-55% range of aa identity shared by the known KS.

The extension of these mutational results to disparate kaurene synthases highlights the importance of the targeted isoleucine in enzymatic cyclization of *ent*-CPP to kaurane type tetracyclic structures, and also may be evolutionarily relevant. In particular, the absolute requirement for gibberellin biosynthesis in all flowering plants (i.e. angiosperms) provides a reservoir of biosynthetic genes, duplication of which (e.g. as a result of the whole genome duplications that are prevalent in plant evolutionary history) has enabled the derivation of alternative diterpenoid metabolism. Accordingly, there are >1,000 known (iso)kaurene derived natural products, as well as ~500 known to be derived from pimaradienes (11). While only a few pimaradiene synthases have been identified (7), none contain an isoleucine at the position identified here, instead containing either threonine or phenylalanine (Figure 4). Hence, it seems likely that the surprisingly facile ability of changes to this particular

isoleucine residue to switch the product outcome of a duplicated KS enzyme to pimaradiene at least partially underlies the observed large numbers of pimaradiene derived natural products. Nevertheless, this primary change in product outcome may then be followed by secondary changes to restore some of the catalytic activity that seems to be lost with this particular mutation and increase specificity for pimaradiene production.

Production of *ent*-pimara-8(14),15-diene (**4**) by mutant (iso)kaurene synthases represents interruption of the normal cyclization of bicyclic *ent*-CPP to tetracyclic (iso)kaurene at the tricyclic pimarenyl stage (Scheme 1). While there has been a previous report of a single tyrosine to phenylalanine mutation in the structurally defined sesquiterpene 5-epi-aristolochene synthase that interrupts a hydrogen bond network and results in the specific production of a similar abortive partially cyclized product (12), this effect appears to be specific to that particular enzyme, revealing an expected stable hydrocarbon intermediate, and is not more broadly applicable (13). There is no evidence that a stable pimaradiene hydrocarbon is formed in (iso)kaurene synthase reactions, and the production of pimaradiene by these mutant diterpene synthases presumably represents deprotonation of the initially formed pimaren-8-yl carbocation (**5**) intermediate.

The production of pimara-8(14),15-diene (**4**) seems to readily occur upon mutation of this particular isoleucine to threonine with the isokaurene synthases but be less favorable in KS. This almost certainly arises from the fact that production of pimara-8(14),15-diene (**4**) results from deprotonation of the same carbon that is deprotonated in formation of isokaur-15-ene (**3**). Thus, I664T mutants of the isokaurene synthases OsKSL5i and OsKSL6 can readily quench pimaren-8-yl$^+$ (**5**) by deprotonation to pimara-8(14),15-diene (**4**). By contrast,

kaurene synthases deprotonate an extracyclic methyl group, which would not quench the pimaren-8-yl$^+$ (**5**) intermediate, and the corresponding isoleucine to threonine mutation is less specific in this context, producing a mixture of pimaradiene with smaller amounts of kaurene. This presumably reflects the lack of an appropriate group for ready deprotonation of the pimaren-8-yl$^+$ (**5**) intermediate, setting up a kinetic competition between such immediate deprotonation to pimara-8(14),15-diene (**4**) and further cyclization that enables deprotonation to kaur-16-ene (**2**) as specified by the original wild type KS.

Given the nature of the mutations studied here, we hypothesize that the polarity introduced by the threonine hydroxyl group may act to stabilize the pimaren-8-yl$^+$ (**5**) intermediate long enough for deprotonation to occur. Consistent with this hypothesis, in modeled structures the relevant residue is located in the F helix on the same side of the active site as the DDXXD Mg^{2+} binding motif that is involved in pyrophosphate positioning, but towards the bottom of the cavity, which can be readily pictured as near the location of the pimaren-8-yl carbocation (Figure 5). The observation of a phenylalanine residue at this position in a pimaradiene synthase (Figure 4) also is consistent with this hypothesis, as the ability of aromatic residues to stabilize carbocations has long been recognized (14). In addition, the ability of the converse threonine to isoleucine mutation of the pimaradiene synthase OsKSL5j to drive secondary cyclization to tetracyclic diterpenes is further consistent with the lack of stabilization of the pimaren-8-yl$^+$ (**5**) intermediate leading to such secondary (tetra)cyclization.

Our hypothesis regarding the role of this particular isoleucine in the secondary (tetra)cyclization mediated by diterpene synthases such as KS is consistent with structural

and mechanistic studies indicating that terpene synthase active sites present an overall hydrophobic environment and largely control product outcome through the conformation/fold of the substrate (i.e. relative positioning of the electrophilic double bonds) (2). However, while selected carbocation intermediates appear to be stabilized within terpene synthase active sites [e.g. ion pairing between the released pyrophosphate anion and beyeran-16-yl carbocation (**6**) intermediate in KS (A. Roy, F.G. Roberts, P.R.W., R.J.P., R.M. Coates; manuscript in preparation)], our hypothesis further implies that product outcome also is directed by the lack of stabilization of other carbocation intermediates (e.g. the pimaren-8-yl$^+$ (**5**) intermediate in KS catalyzed cyclization). These observations may be coupled, as the presence of the ionized pyrophosphate group in the active site should exert a significant electrostatic effect. In particular, driving terpene synthase reactions towards intermediates wherein the carbocation is localized proximal to the multiple counter-ion charges on the pyrophosphate [e.g. in KS from pimaren-8-yl$^+$ (**5**) to beyeran-16-yl$^+$ (**7**)]. This would be consistent with the observation that aza-analogs are generally observed to bind in orientations that provide counter-ion pairing between the aza group and pyrophosphate in tertiary crystal structures, regardless of the expected catalytically productive binding mode (15). Notably, the exertion of such an electrostatic effect on product outcome by the released pyrophosphate moiety also would provide a novel example of substrate-assisted catalytic specificity.

Regardless of the mechanism by which the isoleucine to threonine mutation alters diterpene synthase product outcome, the ability of this single residue change to make such a dramatic difference in product outcome is remarkable. Further, the demonstrated effect of this single residue ‘switch’ in the widely disparate, albeit functionally conserved, kaurene

synthases found in all plants for gibberellin biosynthesis, coupled to sequence analysis of the currently known pimaradiene synthases, suggests that changes to this particular isoleucine are likely to be at least partially responsible for the observed large numbers of pimaradiene derived natural products, potentially providing an example of how secondary metabolism may have been derived from primary (gibberellin) metabolism. Thus, we have followed evolution's lead to a single residue switch for diterpene synthase product outcome that also may have wider evolutionary implications.

MATERIAL AND METHODS

General procedures

Authentic standards for the diterpenes identified here were kindly provided by Dr. Robert Coates. Unless otherwise noted, all other chemicals were purchased from Fisher Scientific (Loughborough, Leicestershire, UK), and molecular biology reagents from Invitrogen (Carlsbad, CA, USA). Gas chromatography (GC) was performed with Agilent (Palo Alto, CA) 6890N GC instruments using an HP-1 column with flame ionization detection (FID) or an HP-5 column with mass spectrometry (MS) detection using a 5973N mass selective detector in electron ionization (70 eV) mode located in the W.M. Keck Metabolomics Research laboratory at Iowa State University. Samples (5 μL) were injected at 40°C in splitless mode, the oven temperature held at 40°C for 3 min., then raised at 20°C/min. to 300°C, and held there for 3 min. MS data were collected from 50 to 500 *m/z* during the temperature ramp and final hold.

Recombinant constructs

The OsKS1, OsKSL5i, and OsKSL6 genes have been previously described (8), while AtKS was previously cloned and kindly provided by Dr. Shinjiro Yamaguchi (16). OsKSL5j was cloned from sub-species *japonica* cultivar Nipponbare rice using the same primers and methods reported for isolation of OsKSL5i (8). All genes were transferred into the Gateway vector system via PCR amplification and directional topoisomerase mediated insertion into pENTR/SD/D-TOPO, and verified by complete sequencing. Site directed mutagenesis was carried out via PCR amplification of the pENTR constructs with overlapping mutagenic primers, and the mutant genes verified by complete sequencing. The resulting wild type and mutant genes were then transferred via directional recombination to the T7-promoter N-terminal GST fusion expression vector pDEST15.

Enzymatic analysis

All constructs were expressed, purified, and the product profile of the resulting recombinant proteins characterized in coupled assays using the *ent*-CPP synthase from maize (An2) to convert (*E,E,E*)-geranygeranyl diphosphate (GGPP) (Sigma-Aldrich, St. Louis, MO) to *ent*-CPP (**1**) as previously described (8). Briefly, 1 mL reactions in assay buffer (50 mM Hepes, pH 7.2, 0.1 mM $MgCl_2$, 10 glycerol, and 5 mM fresh DTT) containing 50 µM GGPP and 100 nM An2 were run ~ 2 hrs at room temperature, which is sufficient to completely convert the GGPP entirely to *ent*-CPP (**1**). Then equivalent volumes of individual purified GST tagged diterpene synthases were added, along with $MgCl_2$ to a final concentration of 10 mM, and the reactions incubated over night at room temperature prior to extraction with hexanes (3 x 1 mL), partial purification over short silica gel columns, and subsequent GC-MS analysis. In addition, each construct was co-expressed in C41 *E. coli*

with a GGPP and *ent*-CPP synthase carried together on a pACYCDuet vector (Novagen, San Diego, CA) and the resulting product isolated as recently described (17). Briefly, the co-transformed *E. coli* were grown and recombinant protein expression induced in mixed phase cultures, specifically liquid media growth cultures containing 2% (wt/vol) adsorbent HP-20 Diaion resin (i.e. liquid and solid phases). The production of diterpene(s) under these conditions is easily assessed by their elution from the HP-20 beads with organic solvent. The resulting diterpene products were identified by GC-MS based comparison to authentic standards, and relative amounts quantified by GC-FID analysis.

Kinetic analysis was carried out using An2 to convert [1-^{3}H]-GGPP (American Radiolabeled Chemicals, St. Louis, MO) to *ent*-CPP (**1**), as described above. Complete conversion was verified by enzymatic dephosphorylation and GC-FID analysis, as previously described (18). This provided 50 μM substrate stock solutions for dilution into subsequent KS reactions. The kinetic assays were performed and analyzed much as previously described (19). Briefly, duplicate 1 mL reactions in assay buffer with 7.5 mM $MgCl_2$ containing 10 nM purified GST-AtKS and 0.1 mg/mL α-casein were initiated by the addition of labeled *ent*-CPP (**1**), run for 1 min. at room temperature, and stopped by the addition of KOH to 0.2 M and EDTA to 15 mM. Although for analysis of the GST-AtKS:I664T mutant the enzymatic concentration was increased to 40 nM and incubation times to 2 min. The produced diterpenes were then extracted, the pooled extract passed over a short silica gel column, production formation assessed by scintillation count, and the resulting data analyzed using Kaleidagraph (Synergy, Reading, PA).

Bioinformatics

All alignments and protein structure manipulation/visualization were performed with the VectorNTI software package (Invitrogen). Modeled protein structures were obtained through the SWISS-MODEL service (20), using alignments of the relevant portions of the target diterpene synthases with 5-epi-aristolochene synthase, whose known structure (21) served as the template.

ACKNOWLEDGEMENTS

We thank Professor Robert M. Coates (University of Illinois) for providing authentic standards, Dr. Shinjiro Yamaguchi (RIKEN) for the AtKS clone, and Sladjana Prisic and Ke Zhou for technical assistance. This work was supported by grants from the NIH (GM076324), NSF (MCB-0416948), and USDA-CSREES-NRI (2005-35318-15477) to R.J.P.

REFERENCES

1. Yoshikuni, Y., Ferrin, T. E., & Keasling, J. D. (2006) *Nature* **440,** 1078-1082.
2. Christianson, D. W. (2006) *Chem. Rev.* **106,** 3412-3442.
3. Greenhagen, B. T., O'Maille, P. E., Noel, J. P., & Chappell, J. (2006) *Proc Natl Acad Sci U S A.* **103,** 9826-9831.
4. Kollner, T. G., Schnee, C., Gershenzon, J., & Degenhardt, J. (2004) *Plant Cell* **16,** 1115-1131.
5. Coates, R. M. & Cavender, P. L. (1980) *J. Am. Chem. Soc.* **102,** 6358-6359.
6. Olszewski, N., Sun, T.-P., & Gubler, F. (2002) *Plant Cell* **Supplement 2002,** S61-S80.

7. Peters, R. J. (2006) *Phytochemistry* **67,** 2307-2317.

8. Xu, M., Wilderman, P. R., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N., Coates, R. M., & Peters, R. J. (2007) *Phytochemistry* **68,** 312-326.

9. Kanno, Y., Otomo, K., Kenmoku, H., Mitsuhashi, W., Yamane, H., Oikawa, H., Toshima, H., Matsuoka, M., Sassa, T., & Toyomasu, T. (2006) *Biosci. Biotechnol. Biochem.* **70,** 1702-1710.

10. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003) *Nat. Biotechnol.* **21,** 796-802.

11. Buckingham, J. (2002) *Dictionary of Natural Products (on-line web edition)* (Chapman & Hall/CRC Press).

12. Rising, K. A., Starks, C. M., Noel, J. P., & Chappell, J. (2000) *J. Am. Chem. Soc.* **122,** 1861-1866.

13. Peters, R. J. & Croteau, R. B. (2002) *Proc. Natl. Acad. Sci. U.S.A.* **99,** 580-584.

14. Dougherty, D. A. (1996) *Science* **271,** 163-168.

15. Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B., & Christianson, D. W. (2002) *Proc Natl Acad Sci U S A.* **99,** 15375-15380.

16. Yamaguchi, S., Sun, T., Kawaide, H., & Kamiya, Y. (1998) *Plant Physiol.* **116,** 1271-1278.

17. Cyr, A., Wilderman, P. R., Determan, M., & Peters, R. J. (2007) *J. Am. Chem. Soc.* **129,** 6684-6685.

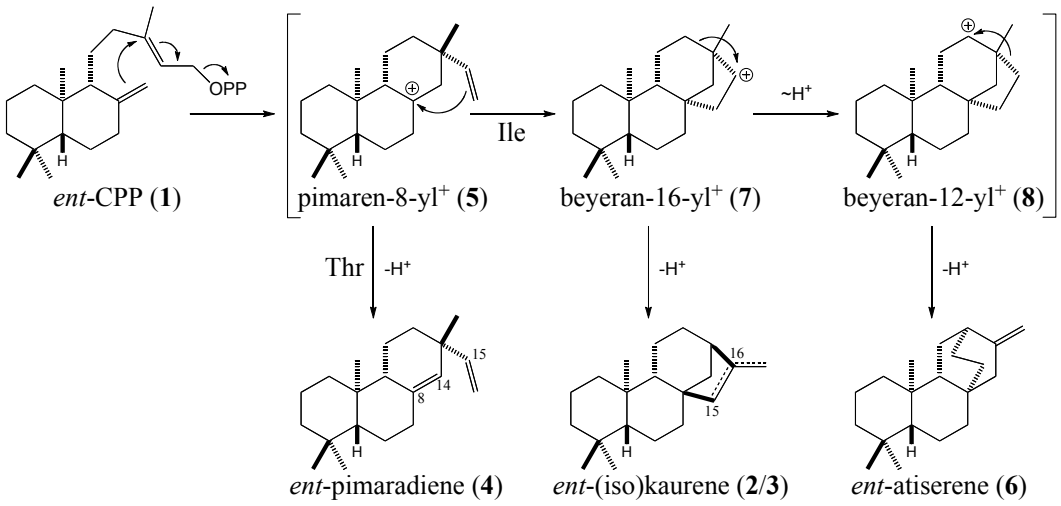
18. Prisic, S., Xu, J., Coates, R. M., & Peters, R. J. (2007) *ChemBioChem* **8,** 869-874.

19. Peters, R. J., Flory, J. E., Jetter, R., Ravn, M. M., Lee, H.-J., Coates, R. M., & Croteau, R. B. (2000) *Biochemistry* **39,** 15592-15602.

20. Guex, N. & Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714-2723.

21. Starks, C. M., Back, K., Chappell, J., & Noel, J. P. (1997) *Science* **277,** 1815-1820.

22. Yamaguchi, S., Saito, T., Abe, H., Yamane, H., Murofushi, N., & Kamiya, Y. (1996) *Plant J* **10,** 101-111.



Scheme 1: Cyclization mechanism for pimaradienes, kaurenes, and atiserene. Diphosphate ionization-initiated cyclization of *ent*-CPP (**2**) to a pimaren-8-yl$^+$ (**5**) intermediate, may be followed by secondary cyclization to a beyeran-16-yl$^+$ (**7**) intermediate that can either undergo ring rearrangement to the kauranyl ring structure, or a 1,3-hydride shift to a beyeran-12-yl$^+$ (**8**) intermediate that undergoes ring rearrangement to the atiseranyl ring structure. In each case the final carbocation intermediate is quenched by deprotonation [dotted bonds indicate alternative double bond placement in *ent*-kaur-16-ene (**2**) versus *ent*-isokaur-15-ene (**3**)].

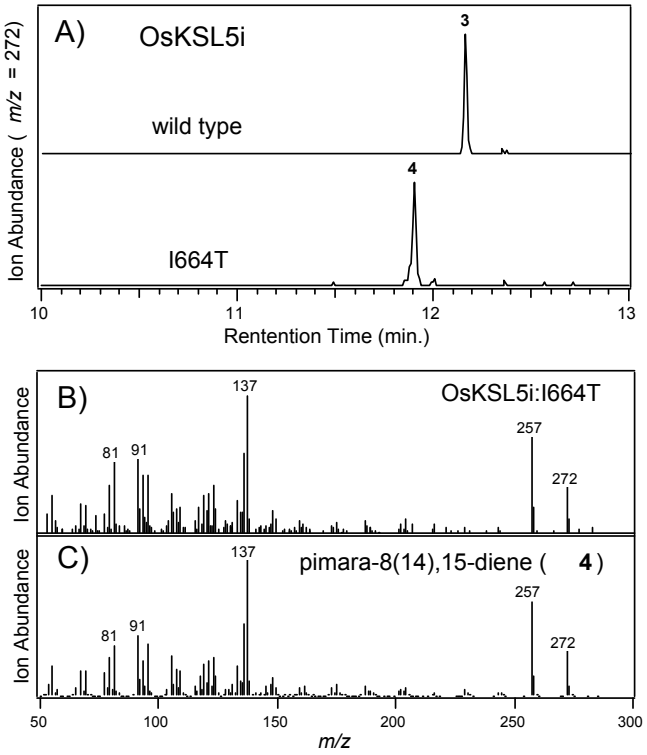


Figure 1: Effect of I664T mutation on OsKSL5i product outcome. A) Chromatograms from GC-MS analysis of wild type and I664T mutant of OsKSL5i (as indicated). Numbers correspond to the chemical structures defined in the text and Scheme 1. B) Mass spectra from the OsKSL5i:I664T product peak (RT = 11.91 min.). C) Mass spectra from an authentic sample of pimara-8(14),15-diene (**4**) (RT = 11.92 min.).

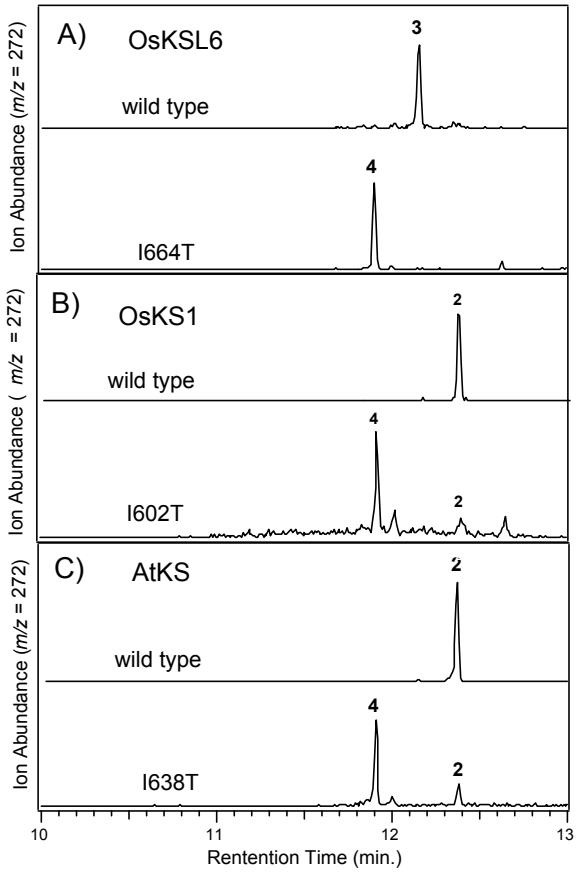


Figure 2: Effect of isoleucine to threonine mutation on (iso)kaurene synthase product outcome. Chromatograms from GC-MS analysis of the indicated diterpene synthases (wild type or indicated mutant). A) OsKSL6. B) OsKS1. C) AtKS. Numbers correspond to the chemical structures defined in the text and Scheme 1. Enzymatic products were identified by comparison of retention time and mass spectra to authentic standards (e.g. see Figure 1).

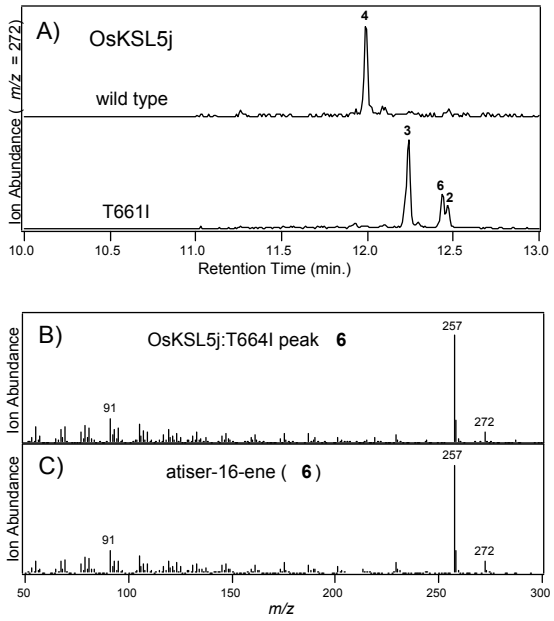


Figure 3: Effect of threonine to isoleucine mutation on OsKSL5j product outcome. A) Chromatograms from GC-MS analysis of wild type and I664T mutant of OsKSL5j (as indicated). Numbers correspond to the chemical structures defined in the text and Scheme 1. B) Mass spectrum from the novel OsKSL5j:T664I product peak labeled **6** (RT = 12.43 min.). C) Mass spectrum from an authentic sample of atiser-16-ene (**6**) (RT = 12.43 min.).

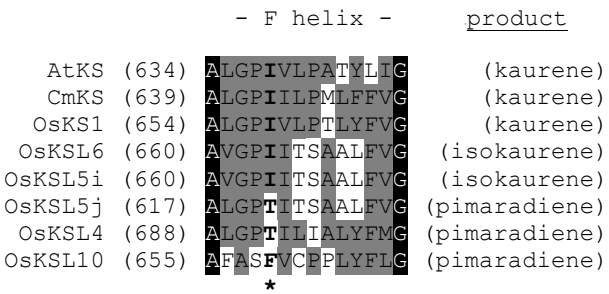


Figure 4: Sequence comparison of the portion of the F helix containing the targeted residue position (indicated in bold and by the *) from selected diterpene synthases whose product structures are indicated on the right. OsKSL nomenclature has been previously defined and followed here (8). CmKS refers to the KS from pumpkin (*Cucurbita maxima*) (22).

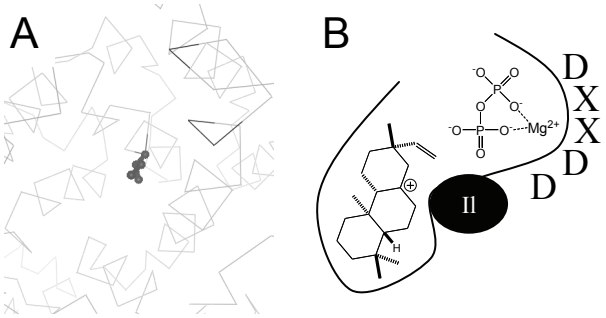


Figure 5: Location of the targeted isoleucine residue. A) In the active site of the modeled structure for AtKS (with the isoleucine side chain depicted in ball-and-stick rendering). Also indicated by darker coloring is the Cα backbone position of the aspartates from the DDXXD Mg^{2+} binding motif. B) Schematic depiction of the isoleucine (Ile) residue and DDXXD motif relative to the ionized diphosphate and pimaren-8-yl carbocation (**5**) intermediate.

Chapter VI. A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase

P. Ross Wilderman and Reuben J. Peters

Abstract

Terpene synthases often mediate the committed step in natural product biosynthetic pathways, so alteration of product outcome is integral in derivation of novel natural products. Upon comparison, we have identified a single residue switch in conifer diterpene synthases similar to that recently reported in rice. This switch of Alanine to Serine effectively changes abietadiene synthase to a pimaradiene synthase. Our results show this switch is not simply limited to the diterpene synthases of grasses.

Main Body

Terpene synthases often catalyze the committed step in natural products biosynthetic pathways and mediate complex reactions, leading to great interest in their enzymatic mechanisms.[1] The resin acid constituents of the oleoresin defensive secretion of grand fir (*Abies grandis*) are derived from a mixture of abietadiene double bond isomers (**1-3**) produced by a single diterpene cyclase, abietadiene synthase (AgAS),[2] while those of Norway spruce (*Picea abies*) arise from two closely related diterpene cyclases that individually produce either isopimara-7,15-diene (**4**) (PaIPS) or abietadienes (**1-3**) (PaLAS).[3] The C13β methyl configuration responsible for the iso-designation of **4** has been shown to be relevant to the pimarenyl^{+} (**5**) intermediates produced en route to abietadienes in the reaction

catalyzed by AgAS,[4] and presumably PaLAS, suggesting that PaIPS terminates its cyclization reaction by deprotonation of a common isopimar-15-en-8-yl$^+$ (**5a**$^+$) intermediate (Scheme 1). In previous work we found that a single residue change was sufficient to 'short circuit' the complex cyclization reaction catalyzed by *ent*-kaurene synthases to instead produce *ent*-pimara-8(14),15-diene, presumably by deprotonation of a mechanistically relevant *ent*-pimar-15-en-8-yl$^+$ intermediate.[5] Here we demonstrate that AgAS product output can be switched from abietadienes to pimaradienes by a similar single residue change, albeit with subtle differences that have important mechanistic implications.

Early work by Wenkert[6] demonstrated the potential biogenetic origin of abietanes from pimaradienes initially suggested by Ruzicka.[7] Later mechanistic work with recombinant AgAS (rAgAS) established the intermediacy of isopimarenyl$^+$ intermediates in enzymatic production of abietadienes from copayl diphosphate (CPP, **6**).[4] In addition, AgAS has been shown to produce sandaracopimara-8(14),15-diene (**7**) as a minor component (~2%) of its product output,[8] suggesting **7** as a potential stable intermediate. However, labeling studies demonstrated intramolecular proton transfer[9] and **7** is not converted to abietadienes by rAgAS,[2] indicating that if **7** is a true intermediate it is only transiently formed en route to abietadienes in the AgAS active site. This intramolecular proton transfer entails a shift from tertiary (**5a**$^+$) to secondary carbocation that is driven, at least in part, by ion pairing of the resulting isopimar-8(14)-en-15-yl$^+$ (**5b**$^+$) with the pyrophosphate anion released by initiating ionization of CPP (**6**).[4]

Recent work has established the dramatic plasticity of terpene synthases, with small numbers of amino acid changes being sufficient to drive significant changes in enzymatic activity (particularly product outcome).[5,10] However, previous alanine scanning mutagenesis of all the polar and charged residues in the relevant modeled active site of rAgAS generally led to very limited changes in product profile, with the only exceptions arising from mutant enzymes that were also severely kinetically compromised (>10,000-fold reductions in

catalytic efficiency).[11] Note that AgAS, as well as PaIPS and PaLAS, are all bifunctional diterpene cyclases that catalyze both mechanistically unusual protonation-initiated (i.e. class II) and the more typical diphosphate ionization-initiated (i.e. class I) cyclization reactions in separate active sites,[12] and the results presented here exclusively pertain to the class I active site analogous to that found in most terpene synthases.

In a previous report we demonstrated that substitution of Thr for a specific Ile conserved in all the known *ent*-kaurene synthases was sufficient to switch the product profile of these diterpene cyclases to largely (≥80%) *ent*-pimara-8(14),15-diene.[5] Given the mechanistic similarity between these results and the production of isopimaradiene (**4**) by PaIPS, which is quite closely related to diterpene cyclases that produce abietadienes (e.g. PaIPS and AgAS are 86% identical), we hypothesized that there might be a similar single residue switch responsible for this difference in product outcome. Alignment of these conifer diterpene cyclases with those from rice (*Oryza sativa*) utilized in our previous study revealed that there was only a small change in aliphatic side chain size at the corresponding position, i.e. Val727 in AgAS, which is also Val in PaLAS, but Leu in PaIPS (Figure 1). However, this type of change in *ent*-kaurene synthases did not alter product output (e.g. Val substitution for the relevant Ile; data not shown), indicating that introduction of polarity into the active site is critical. Intriguingly, four residues away there is a Ser found in PaIPS in place of an Ala that is conserved in the abietadienes specific diterpene cyclases. This residue was located in the modeled active site of AgAS, specifically one turn of the relevant helix away from the single residue switch found for the *ent*-kaurene synthases. Considering the difference in configuration of the CPP (**6**) substrate utilized by the conifer diterpene cyclases and the *enantio*-stereoisomer utilized by *ent*-kaurene synthases, we reasoned that this residue might correspond to the hypothetical single residue switch.

Substitution of this Ala by Ser in rAgAS (rAgAS:A723S) results in a diterpene cyclase that specifically produces pimaradienes, specifically 75% isopimara-7,15-diene (**4**)

and 21% sandaracopimara-8(14),15-diene (**7**), with only minor amounts (~4%) of abieta-7,13-diene (**1**) (Figure 2).[13] Notably, both **4** and **7** have the C13β methyl stereochemistry expected from the known involvement of isopimarenyl$^+$ intermediates in cyclization to abietadienes. In addition, rAgAS:A723S is kinetically comparable to the wild type enzyme, actually exhibiting slightly increased specific activity (~2-fold), consistent with the possibility that this single residue switch might be evolutionarily relevant (i.e. mediate sufficient metabolic flux to maintain resin acid production). While PaIPS produces exclusively **4** and other changes would be necessary to convert AgAS to such a selective isopimaradiene synthase, the ability of this single residue change to convert rAgAS from producing >95% abietadienes (**1-3**) to >95% pimaradienes (**4** and **7**) is nevertheless quite remarkable.

The increased active site volume resulting from the Thr for Ile mutation in our previous study left open the possibility that the introduced hydroxyl group might coordinate a novel active site water that directly deprotonated the relevant pimarenyl$^+$ intermediate. However, the Ser for Ala mutant reported here reinforces our hypothesis that the introduced hydroxyl group stabilizes the pimar-15-en-8-yl$^+$ intermediate, albeit with the necessary shift in active site position presumably dictated by the change in CPP stereochemistry, for deprotonation by other means (note that use of the hydroxyl side chain as a general base seems implausible given the generally hydrophobic nature of terpene synthase active sites, including that modeled for rAgAS). Indeed, production of isopimar-7,15-diene (**4**) results from proton removal from C7, which is similarly deprotonated in the production of abieta-7,13-diene (**1**), and is the only position that is deprotonated in the final abieta-8(14)-en-13-yl$^+$ (**8**$^+$) intermediate to form the normally observed abietadienes (**1-3**) that also will quench the presumed common pimar-15-en-8-yl$^+$ (**5a**$^+$) intermediate (Scheme 1), indicating use of a pre-existing general base in the active site of rAgAS:A723S. This further implies that sandaracopimaradiene (**7**) is not a true intermediate in production of the abietadienes, as the

presence of an active site general base for even transient production of this 8(14),15-diene would then be expected to result in dominant production of **7**, rather than the relatively small proportion actually observed.

In both the abietadienes and kaurenes reaction mechanisms transitions from tertiary pimar-15-en-8-yl carbocations to subsequent secondary carbocation intermediates, pimar-8(14)-en-15-yl$^+$ (**5b**$^+$) in the case of the abietadienes, are driven by ionic pairing with the pyrophosphate group released upon ionization of the CPP substrate (i.e. the secondary carbocations are significantly closer to the pyrophosphate than the preceding tertiary carbocations; Scheme 1).[4,14] Accordingly, we hypothesize that the introduced hydroxyl group, at the relevant position in either rAgAS or the previously examined *ent*-kaurene synthases, simply acts to stabilize pimar-15-en-8-yl$^+$ intermediates long enough for deprotonation to occur, and that the more extended reactions leading to production of abietadienes or kaurene products are under kinetic control, with the ionized pyrophosphate group assuming a critical role in directing product outcome, providing now two examples of substrate assisted catalytic specificity. Such a role for the ionized pyrophosphate group may then be a general feature of the cyclization/rearrangement reactions catalyzed by terpene synthases.

Regardless of the exact means by which these mutations alter product outcome, their ability to do so is striking. Further, particularly given the previously demonstrated conversion of an *ent*-pimaradiene synthase to the production of *ent*-kaurene by the converse Thr to Ile mutation,[5] the finding of a similar single residue switch for alternative CPP stereochemistry reported here offers the possibility of using such changes in engineering product outcome in diterpene synthases more generally.

Acknowledgements

We thank Professors Robert M. Coates (Univ. Illinois) and Rodney B. Croteau (Washington State Univ.) for authentic standards, and Meimei Xu and Ke Zhou for technical assistance. This work was funded by a grant from the NIH (GM076324) to R.J.P.

References

(1) Christianson, D. W. *Chem. Rev.* 2006, *106*, 3412-3442.

(2) Peters, R. J.; Flory, J. E.; Jetter, R.; Ravn, M. M.; Lee, H.-J.; Coates, R. M.; Croteau, R. B. *Biochemistry* 2000, *39*, 15592-15602.

(3) Martin, D. M.; Faldt, J.; Bohlmann, J. *Plant Physiol.* 2004, *135*, 1908-1927.

(4) Ravn, M. M.; Peters, R. J.; Coates, R. M.; Croteau, R. B. *J. Am. Chem. Soc.* 2002, *124*, 6998-7006.

(5) Xu, M.; Wilderman, P. R.; Peters, R. J. *Proc. Natl. Acad. Sci. U.S.A.* 2007, *104*, 7397-7401.

(6) Wenkert, E.; Chamberlin, J. W. *J. Am. Chem. Soc.* 1959, *81*, 688-693.

(7) Ruzicka, L.; Eschenmoser, A.; Heusser, H. *Experentia* 1953, *IX*, 357-367.

(8) LaFever, R. E.; Stofer Vogel, B.; Croteau, R. *Arch. Biochem. Biophys.* 1994, *131*, 139-149.

(9) Ravn, M. M.; Coates, R. M.; Jetter, R.; Croteau, R. *J. Chem. Soc. Chem. Commun.* 1998, *1998*, 21-22.

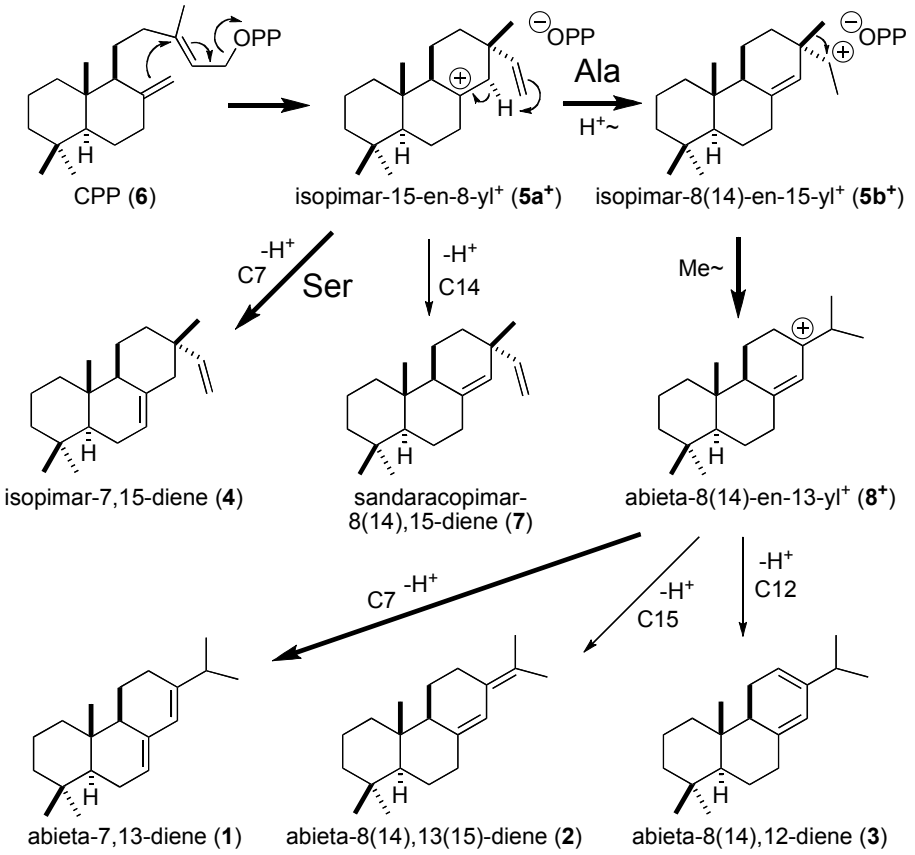
(10) (a) Kollner, T. G.; Schnee, C.; Gershenzon, J.; Degenhardt, J. *Plant Cell* 2004, *16*, 1115-1131. (b) Greenhagen, B. T.; O'Maille, P. E.; Noel, J. P.; Chappell, J. *Proc Natl Acad Sci U S A.* 2006, *103*, 9826-9831. (c) Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D. *Nature* 2006, *440*, 1078-1082. (d) Yoshikuni, Y.; Martin, V. J. J.; Ferrin, T. E.; Keasling, J. D. *Chem. Biol.* 2006, *13*, 91-98.

(11) Peters, R. J.; Croteau, R. B. *Proc. Natl. Acad. Sci. U.S.A.* 2002, *99*, 580-584.

(12) Peters, R. J.; Ravn, M. M.; Coates, R. M.; Croteau, R. B. *J. Am. Chem. Soc.* 2001, *123*, 8974-8978.

(13) Enzymatic characterization was carried out as previously described.[2,12]

(14) Roy, A.; Roberts, F.; Wilderman, P.R.; Peters, R.J.; Coates, R.M. *J. Am. Chem. Soc.* 2007, *in press*.



Scheme I. Cyclization to pimaradienes and abietadienes by AgAS.

Synthase		Alignment	Product
AgAS	(717)	NASVSIALGTVVLISALFTG	abietadienes
PaLAS	(707)	NASVSIALGTVVLISALFTG	abietadienes
PaIPS	(715)	NASVSISLGTLVLISVLFTG	isopimaradiene
		* #	
OsKSL5	(652)	AAVDSFALGPTITSAALFVG	*ent*-pimaradiene
OsKSL6	(654)	AAVDSFAVGPIITSAALFVG	*ent*-kaur-15-ene
OsKS1	(648)	NAVVSFALGPIVLPTLYFVG	*ent*-kaur-16-ene

Figure 1. Diterpene synthase alignment. The location of the single residue switch reported here is indicated by * and that previously reported by #.

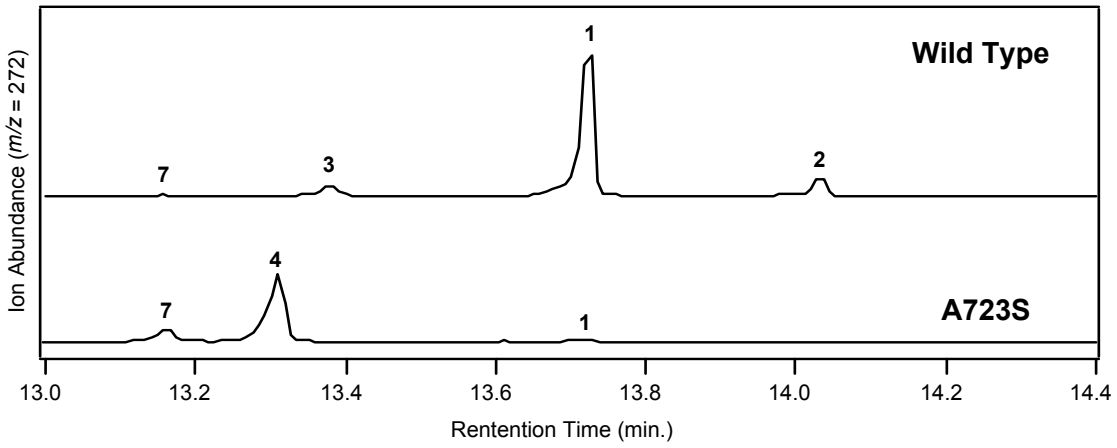


Figure 2. Effect of A723S mutation on rAgAS product outcome. Chromatograms from GC-MS analysis of rAgAS (either wild type or A723S mutant, as indicated). Numbers correspond to the chemical structures defined in the text and Scheme 1, as identified by comparison of retention time and mass spectra to authentic standards.

Chapter VII. General Conclusions and Future Work

Kaurene synthase like enzymes in rice

Building upon the rice (*O. sativa*) genome sequencing projects, we functionally identified a *syn*-pimara-7,15-diene synthase, OsKSL4, the first identified TPS with this substrate specificity [1]. This biosynthetic reaction is the committed step in momilactone biosynthesis. This gene was shown to be up regulated by ultraviolet light and methyl jasmonate, upholding the suggestion previously proposed that plant secondary metabolism is most often regulated at the level of transcription [2]. Further, upon examination of where the previously identified OsCPS4 and OsKSL4 are located in the rice genome, it was found that they were in close physical proximity in the genome, suggesting functionally clustering of TPS genes in common biosynthetic pathways [1,3].

Further work led to the identification of eight kaurene synthase-like genes, two pseudogenes, and one partial gene in rice (*Oryza sativa* ssp. indica) [4]. The enzymes identified produced the known products produced by rice as identified by Mohan and coworkers (*ent*-kaur-16-ene, *ent*-cassa-12,15-diene, *ent*-sandaracopimaradiene, *syn*-pimara-7,15-diene, and *syn*-stemar-13-ene), as well as *ent*-isokaurene (*ent*-kaur-15-ene) and *syn*-stemod-13(17)-ene (Figure 1) [4]. At the time this work was accomplished, another group published a paper describing similar results using a different rice cultivar (*Oryza sativa* ssp. japonica), however one of the enzymes described in both manuscripts produced different products, OsKSL5: *ent*-isokaurene in indica and *ent*-pimara-8(14),15-diene in japonica[4,5]. These sub-species orthologs provided a platform for examination of determinants of product profile in kaurene synthase-like enzymes.

Single residue switch

Prior to this type of examination, a method of producing detectable amounts of products from TPS enzymes needed to be created. This led to the use of the metabolic engineering system that was designed in our laboratory [6]. This system utilized multiple vectors containing at a minimum geranylgeranyl diphosphate synthase and generally a copalyl diphosphate synthase of the desired stereochemistry, as well. This setup enables biosynthesis of significant amounts of product and provides a quick means of examining perturbations in product profiles of enzymes of interest.

Utilizing this metabolic engineering system, product production was examined in kaurene synthase-like enzymes. As isoprenoids have a wide range of chemical structures and biological activities, determining how products are formed will lead to understanding how to utilize the potential of these natural products. With the advances in understanding kaurene synthase-like enzymes in rice and having the metabolic engineering system in our laboratory, a set of experiments examining product determinants was conceived.

OsKSL5 has been reported as an *ent*-pimaradiene synthase in O. sativa ssp. Japonica, but it was shown to be an isokaurene synthase in O. sativa ssp. Indica [4,5]. This enzyme is >98% identical between the two cultivars, however there appear to be three amino acids in or very near the active site which are different between the two enzymes (Fig. 2). The I664T mutant in OsKSL5i produced *ent*-pimaradiene, and this mutation was found to switch product profile in OsKSL1 and AtKS, kaurene synthases from rice and *Arabidopsis*, respectively [7]. With the substitution of a polar residue into the place of a hydrophobic residue, it is possible the polar nature of threonine stabilizes a carbocation intermediate allowing for a truncated cyclization cascade. With this knowledge, we proposed this single residue switch was not limited to the kaurene synthase-like enzymes in rice.

Examination of the abietane synthases from conifers showed a similar single residue substitution in the amino acids that mapped to the active site using macromolecular

modeling. This was a switch from an alanine in abietadiene synthases from grand fir, Norway spruce, and ginkgo biloba to a serine in an isopimaradiene synthase from Norway spruce. By mutating recombinant abietadiene synthase from grand fir (*Abies grandis*) at this location (rAgAS:A723S), the product profile of the enzyme switched from >95% abietanes to 95% pimaradienes. This is another example of a single residue switch utilized by evolution.

Future Experiments

In addition to the information provided in this thesis forms the basis from which multiple experiments have been conceived that will be greatly assisted by the presented work. An interesting portion of diterpene synthase active sites is the proposed secondary metal binding motif [(N/D)Dxx(T/S)xxxE] (Fig. 3) [8]. On the basis of the 5-EAS precedent, the first and last aspartates (D621 and D625) in the canonical DDxxD motif of this second active site of AgAS chelate a magnesium ion (Mg_a^{2+}) along with the assistance of the highly conserved E699 [9]. A second magnesium ion (Mg_b^{2+}) is bound by N765 (almost always an aspartate in other synthases), a nonconserved polar residue (in this case, T769), and the highly conserved E773. On substrate binding, a third magnesium ion (Mg_c^{2+}) binds to the diphosphate moiety and is further chelated by D621 (which then ligates both Mg_a^{2+} and Mg_c^{2+}). In AgAS, the polar residue in this motif has been mutated to alanine (T769A), which creates a severely impaired enzyme [10]. The addition of alanine to this area prevents the stabilizing influence of the polar residue. The effect of a mutation of the polar residue to glycine, where water could interact through hydrogen bonding with the glycine proton and Mg_b^{2+}, would be interesting. The Ala substitution will reduce enzyme activity, as shown previously, but the Gly mutation may still be active [10]. This is due to the Gly residue being present in naturally occurring enzymes including AtKS, OsKS1, and OsKSL7. Presumably a water molecule should occupy the space generated upon removal of the Thr

side chain but would by occluded by Ala. Kinetic analysis will be completed with the method used previously with AgAS [10].

In addition to these mechanistic experiments, further investigations into substrate and product specificity have been envisioned. Using a combination of macromolecular modeling and sequence conservation of DTS enzymes, direct mutational analysis of the residues that impose the observed substrate and product specificity should be carried out. Terpene synthase active sites are found in the central cavity of a helical bundle domain of known structure where the catalytically required divalent metal ions occupy an invariant position, defined by a highly conserved DDxxD motif, at the top of the active site [9]. In turn, this fixes the position of the substrate pyrophosphate moiety. Upon further examination, the *ent*- and *syn*-isomers of CPP are remarkably similar with the same bicyclic ring structure and an isoprenyl moiety attached to the pyrophosphate. The only difference in their structures is the position of the C10 methyl group, 10α-positioning for *ent*-CPP and 109β-positioning for *syn*-CPP. Therefore, differences in stereochemistry and relative positioning of the bicyclic ring must be dictated by changes lower in the active site cavity, defining an area of interest for selection of residues for mutational analysis.

Such analysis has been already applied to the known DTS enzymes to identify a number of residues that are hypothesized to impact the shape of the active site cavity and affect substrate and product specificity. For example, the modeled active site cavity in abietane synthases (specific for CPP of normal stereochemistry) has a distinctly different shape than that of the ent-CPP specific kaurene synthases. In particular, a conserved Tyr residue underneath the DDxxD motif and an opposing Ile affects the shape of abietane synthase active sites (Fig. 4). In contrast, *ent*-kaurene synthases contain Met or Leu in place of the Tyr and have an opposing conserved Phe rather than Ile (Fig. 4). Manual docking further predicts steric clash between these residues and the unreactive stereoisomers of CPP. Therefore, these, and other similarly identified residues, have been reciprocally mutated to

the residue(s) found in synthases with different specificity. Abietadiene synthase mutants currently available through site directed mutagenesis done in our lab by Ke Zhou include Y696M, Y696L, and I722F single mutants and Y696M/I722F, and Y696L/I722M double mutants. The addition of aromatic bulk with Phe in the active site of abietadiene synthase should disallow substrate binding and subsequent catalysis; conversely, removal of the bulk of Tyr should open the active site to other stereoisomers of CPP and/or chemistries not observed in wild type enzyme reactions. These mutants should shed some light on enzymatic determinants of substrate specificity in diterpene synthases.

References

1. Wilderman, P. R., Xu, M., Jin, Y., Coates, R. M., Peters, R. J. Identification of *syn*-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiol*., 2004. 135: 2098-2105.
2. Peters, R.J. and Croteau, R.B. 2004. Metabolic Engineering of Plant Secondary Metabolism. *In* G Kishore, ed, Handbook of Plant Biotechnology: Applications of Plant Biotechnology in Agriculture, the Pharmaceutical Industry, and Other Industries, Vol 2. John Wiley & Sons Ltd, London, pp 609-628.
3. Xu, M., Hillwig, M. L., Prisic, S., Coates, R. M., Peters, R. J. Functional identification of rice *syn*-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products. *Plant J,*. 2004. 39: 309-318.
4. Xu, M., Wilderman, P. R., Morrone, D., Xu, J., Roy, A., Margis-Pinheiro, M., Upadhyaya, N., Coates, R. M., & Peters, R. J. Functional Characterization of the Rice Kaurene Synthase-Like Gene Family. *Phytochemistry,* 2007. 68: 312-326.

5. Kanno, Y., Otomo, K., Kenmoku, H., Mitsuhashi, W., Yamane, H., Oikawa, H., Toshima, H., Matsuoka, M., Sassa, T., & Toyomasu, T. Characterization of a rice gene family encoding type-A diterpene cyclases. *Biosci. Biotechnol. Biochem*. 2006. 70: 1702-1710.
6. Cyr, A., Wilderman, P.R., Determan, M., Peters, R.J. A modular approach for facile biosynthesis of labdane-related diterpenes. *J Am Chem Soc*, 2007. 129: 6684-6685.
7. Xu, M., Wilderman, P.R., Peters, R.J. Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc Natl Acad Sci U S A*, 2007. 104: 7397-7401.
8. Rynkiewicz MJ, Cane DE, Christianson DW. Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. *Proc Natl Acad Sci U S A*, 2001. 98:13543-13548.
9. Starks, C.M., Back, K., Chappell, J., Noel, J.P. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science*, 1997. 277: 1815-1820.
10. Peters, R.J. and R.B. Croteau, Abietadiene synthase catalysis: Mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. *Proc Natl Acad Sci, U.S.A*., 2002. 99: 580-584.

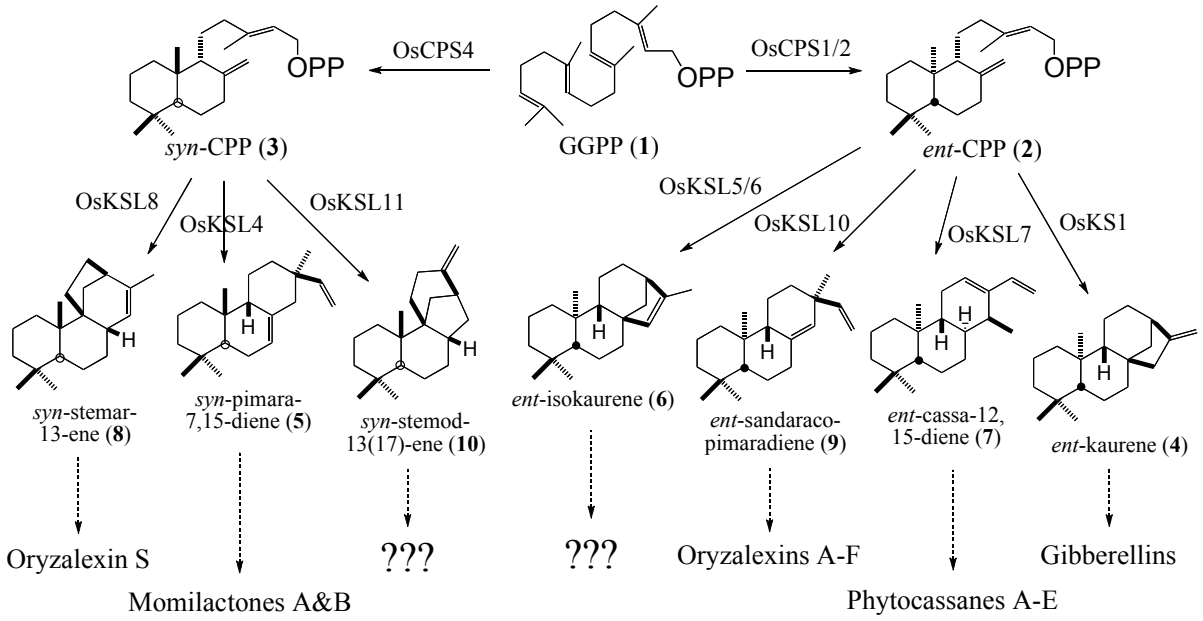


Figure 1. Known labdane-related diterpene cyclization reactions in rice. The corresponding cyclases are indicated, along with their products and, where known, the derived natural products (dashed arrows indicate multiple biosynthetic steps).

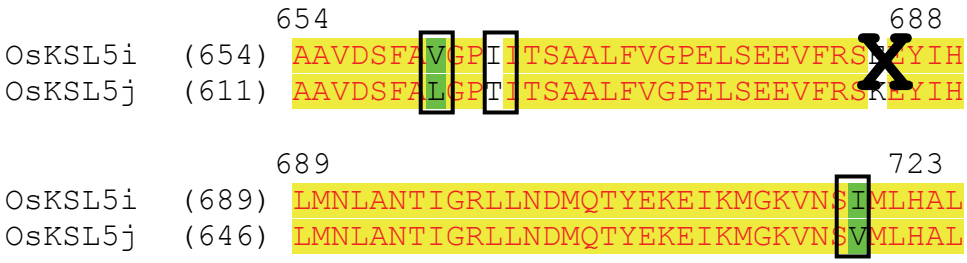


Figure 2. Alignment of OsKSL5i and OsKSL5j showing residues that are different and presumably found in the active site.

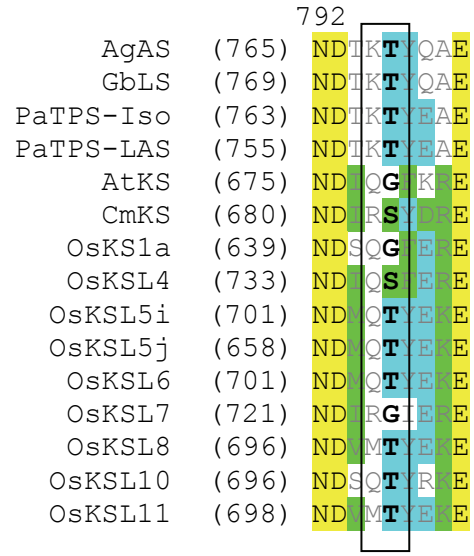


Figure 3. Alignment showing secondary metal binding motif [(N/D)Dxx(T/S)xxxE].

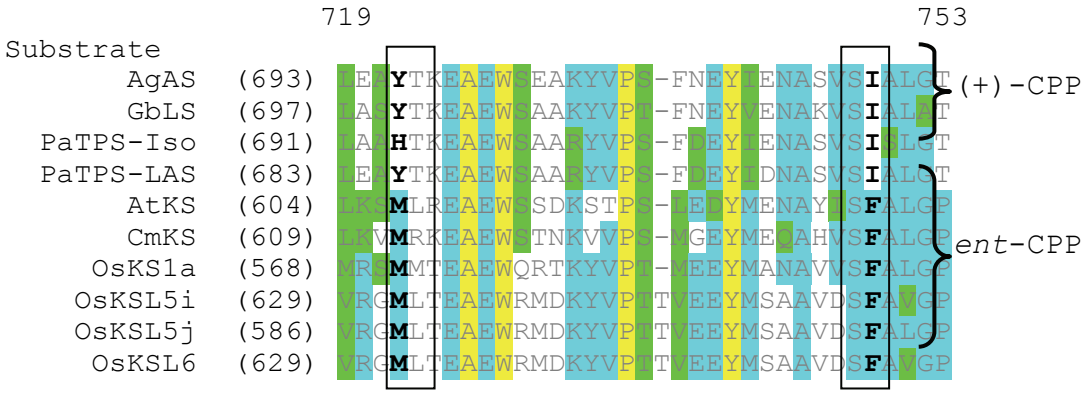


Figure 4. Alignment showing residues of interest in diterpene synthases specific for CPP of normal or *ent*-stereochemistry.

VITA

NAME OF AUTHOR: Paul Ross Wilderman

DATE AND PLACE OF BIRTH: May 7, 1976, Englewood, CO.

DEGREES AWARDED:

B.S in Molecular Biology, Minor in Chemistry, University of Wyoming, 1999

PROFESSIONAL EXPERIENCE:

Undergraduate Research Assistant, Department of Molecular Biology, University of Wyoming, 1998-1999.

Research Assistant, Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, 2003-2007.

PUBLICATIONS:

Wilderman, P. R., M. Xu, et al. (2004). "Identification of Syn-Pimara-7,15-Diene Synthase Reveals Functional Clustering of Terpene Synthases Involved in Rice Phytoalexin/Allelochemical Biosynthesis." Plant Physiol. 135(4): 2098-2105.

Prisic, S., M. Xu, et al. (2004). "Rice Contains Two Disparate ent-Copalyl Diphosphate Synthases with Distinct Metabolic Functions." Plant Physiol. 136(4): 4228-4236.

Xu, M., P. Ross Wilderman, et al. (2007). "Functional characterization of the rice kaurene synthase-like gene family." Phytochemistry. 68(3): 312-326.

Xu, M., P. R. Wilderman, R. J. Peters. (2007). "Following evolution's lead to a single residue switch for diterpene synthase product outcome." Proc Natl Acad Sci U S A. 104(18): 7397-7401.

Cyr, A., P. R. Wilderman, et al. (2007). "A modular approach for combinatorial biosynthesis of labdane-related diterpenes." J Am Chem Soc. 129(21): 6684-6685.

Roy, A., et al. (2007). "Azabeyerane and Azatrachylobane: Potent Mechanism-based Inhibitors of Recombinant Kaurene Synthase from *Arabadopsis thaliana.*" J Am Chem Soc. *In press.*

Wilderman, P. R., and R. J. Peters. (2007). "Abietadiene synthase contains a single residue switch which determines product outcome." J Am Chem Soc. *In review*.